THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: John Arthur Ridgeway

TITLE OF THESIS: Cost Relationships and Capacity Allocation Within the Canadian Railway Industry

DEGREE FOR WHICH THESIS WAS PRESENTED: Master of Arts

YEAR THIS DEGREE GRANTED: 1980

THE UNIVERSITY OF ALBERTA

COST RELATIONSHIPS AND CAPACITY ALLOCATION WITHIN THE CANADIAN RAILWAY INDUSTRY

by

© JOHN ARTHUR RIDGEWAY

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Arts

Department of Economics

EDMONTON, ALBERTA

FALL 1980

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "Cost Relationships and Capacity Allocation Within the Canadian Railway Industry" submitted by John Arthur Ridgeway in partial fulfilment of the requirements for the degree of Master of Arts.

DEDICATION

I dedicate this work to my mother and father whose love taught me the most important lessons in life and to Lois whose simple joy in living shall help me through the years ahead.

ABSTRACT

Studies of the Canadian railway system have become regular events but few have used the tools of statistical cost analysis and production theory. The first chapter reviews some studies using these techniques. Production theory in its dual aspects of profit maximization or cost minimization, as presented by Shephard (1970), is discussed and related to the railway industry in Chapter Two. The chapter concludes with a presentation of Shephard's Lemma which is a very powerful tool for model estimation. Functional forms used in most previous cost studies have required the assumptions of nonjointness, separability and homogeneity. The implications of these assumptions are discussed at the beginning of Chapter Three. The translog function, one of a series of recently developed functional forms not requiring these assumptions, is presented as the model to be used in this study. The chapter is concluded by a discussion of the problems of measuring returns to scale and incremental costs when multiple outputs are produced. Chapter Four specifies a model of Canadian railways using a translog cost function with two inputs and four outputs. An elasticity of transformation measure parallel to Allen's partial elasticity of substitution is developed at the end of Chapter Four. The estimation results are presented in Chapter Five with measures of

returns to scale and incremental costs. Elasticities of supply and demand and Allen's partial elasticities of substitution and transformation are also presented. The final chapter discusses the results as they might affect current concerns in railway policy and suggests areas for further study.

ACKNOWLEDGEMENT

I would like to acknowledge the many hours of patient guidance given me by David Gillen, Mel MacMillan and Terry Levesque during the preparation of this thesis. The typing was supplied by Lynne Rowen whose fine workmanship is clearly evident throughout the thesis. Finally, I would like to thank Lori for her support and friendship and the many insights she has given me both into economics and life in general.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

1.1 Introduction

Railway policy has been a current issue throughout the history of this country since confederation. The railway system was formed to bring Canada into existence and true to its political parentage it has provided over a century of debate and controversy. Many studies, inquiries and commissions have examined the railway problem, most notable are the Royal Commissions which have come into being on average every ten years. Most recently the Snavely and Hall Commissions have examined the railway system in regard to its grain handling operations.

A great deal of the study and discussion of railway policy has concerned the transportation of goods to and from the large land-locked area of western Canada. Rail transportation provided the only efficient means of transportation to and from the prairies until after the second World War. While it no longer holds this position for commodities in general, there are still large movements of goods which can be most efficiently moved by rail. Thus the railway remains a key factor in the prairie economy.

In 1973 the four western provinces put forward, at the Western Economic Opportunities Conference, a proposal for changes to the transportation policy of Canada. Central to this proposal were major changes in the structure of the railway industry. The western premiers sought a revision of the National Transportation Act to emphasize regional development. More specifically they asked that a "Western Transportation Evaluation Authority" be established and that it report to a "Western Canadian Transportation Policy Committee."

Direct changes in the ownership of some railway plant was also proposed. The western provinces also asked that the federal government take over ownership and operation of the roadbed and right of way and to provide access to any public carrier wishing to use it. It was hoped that this would lead the railway industry into more inter-firm competition such as with trucking companies which operate over the same roads. Furthermore, it was argued, the take-over of the railbed would make the transportation system more equitable in regard to the costs born directly by the users of the system. In support of this argument, it was noted that the federal government contributed about 80 percent of the fixed costs for air and water transportation but only about 20 percent for rail. Interestingly, the Hall Commission has since made recommendations that would have the federal government move in the direction of the above proposals.

Full cost disclosure for all forms of transportation was also called for by the western premiers. For many years the west has felt that the pricing policies of the railways have stood as a hindrance to economic development in the region. A lack of cost information has led to many long hours of debate over the years as to the appropriateness of the rate structure and levels of the railway system. No single problem has dominated the discussion of railway policy as much as costing.

That costs of service should be the basis for any rates has long been argued, both from an equity and efficiency point of view. Economic theory is quite explicit on the need for prices to be based on, if not equal to, the marginal costs of production. Ideally all prices should equal marginal costs of production but some situations may arise in which this is not possible or desirable, the simplest case being that of increasing returns to scale. In the early years of the railway it seems almost certain that increasing returns were available, thus any attempt at marginal cost pricing would leave total revenues less than total costs. This does not mean that the marginal cost rule is to be ignored, rather prices should deviate from marginal costs in a systematic manner to attain the most efficient rate structure.[1]

---

[1]For a discussion of the optimal departure from marginal cost pricing see Baumol and Bradford (1970).

One of the assumptions underlying the marginal cost pricing rule is that the distribution of income is optimal. Should this not be the case then a movement away from marginal cost pricing may help to attain a more optimal income distribution. Certain railway policies of the federal government are explicitly designed to redistribute income among the regions of Canada. Examples of such policies are the statutory grain rates, the maritime freight rate subsidy and the 'bridge' traffic rate.

An examination of both the Snavely and Hall Commissions reports reveals that the greatest part of their time and effort was spent on the question of costs. The Canadian Transport Commission is currently involved in a large study of railway costing, further indicating the need for information in this area. Common to these studies is the limited reference to the economic literature on railway costs and costing in general. In response to this the following section reviews a number of these studies.

Following the review of the literature this thesis will proceed to apply received production theory and statistical cost analysis to the problem of railway costs in Canada. A model of the cost relations within Canadian railways will be specified after a discussion of production theory and recent developments in functional forms. Upon estimation of the model, the results will be discussed in light of current railway policy.

1.2 Review of Previous Studies

Statistical estimation of cost functions for railways has not generated a large literature and that which is to be found concerns mostly American railways. The following section briefly reviews the progress of statistical railway cost studies from the late 1940's to the most recent developments upon which this study is based.

Two early studies of railways were undertaken by Borts (1952) and (1960). In his first paper Borts divided the railway operations into line haul and switching operations. For each operation outputs were taken to be exogenous, due to the common carrier nature of the railways and inputs were divided as to those that were endogenous and those that were exogenous. A production function was specified, wherein each of the endogenous inputs was taken to be a function of the outputs and the exogenous inputs. The production relation took the form

$$
\begin{aligned}
X_1 &= f_1(Z_1, Z_2, \ldots, Z_n) \\
X_2 &= f_2(Z_1, Z_2, \ldots, Z_n) \\
&\vdots \\
X_m &= f_m(Z_1, Z_2, \ldots, Z_n)
\end{aligned}
\tag{1.1}
$$

where the X's are the endogenous inputs and the Z's are the exogenous variables.

Endogenous variables for both the line haul and switching operations were, man-hours of labor services, tons of fuel consumed, flow of equipment services measured as expenditure on maintenance and flow of track and structure services, measured as expenditure on maintenance. Line-haul outputs were taken to be loaded freight-car miles, carloads of freight and empty freight car miles. Total tractive capacity of freight locomotives and miles of mainline track represented the exogenous inputs for the linehaul operation. Switching outputs consisted of yard switching locomotive miles, yard switching locomotive hours and car loads of freight. Exogenous inputs were miles of yard switching track, total tractive capacity of yard locomotives and the average number of freight cars standing on the line.

Cross-sectional data from Class I railways in the United States for the year 1948 were used to estimate the model. Two specifications of the model were estimated. The first model used a simple linear specification while the second model divided all variables by a measure of capacity for the railway being observed. Capacity was measured for the line haul by taking the costs of reproduction of all capital equipment employed by the firm and by miles of track devoted to yard switching for the switching operation.

Borts was primarily interested in the returns to scale in the railway industry and calculated the elasti-

city of variable cost to examine this question. He found that for the switching operation the cost elasticity calculated for each model was not significantly different from one, thus indicating constant returns to scale. However, the two models gave quite different results concerning the line hall operation. The first model again gave a cost elasticity not significantly different from one but the second model produced a cost elasticity of slightly greater than one-half thus indicating increasing returns to scale. Borts makes no attempt to choose between the two models.

In his second article, Borts argues that while previous studies of railways have noted that the size of the firm and its location in the country will affect its cost parameters, the specifications used had led to biased results. Borts demonstrates that by incorrectly entering the size of the firm into the estimation of the cost function and due to the 'regression fallacy' previous estimates of returns to scale had been biased toward increasing returns. The regression fallacy is due to actual output deviating from planned output with a lag in expenditure adjustment. Measuring costs by expenditures as is generally done in cost studies means that if output is below the expected level, expenditure will be above the actual cost curves due to excess factors being employed. Conversely, if output is above the expected level then production is facilitated by a reduction in stocks through a lag in expenditure on replacement and maintenance. This

will give observations on expenditure and output that will fall below the true cost function, leading to slope estimates lower than the true slopes of both the short-run and long-run cost functions.

Borts then proceeds to develop and estimate a model for freight services again using cross-sectional data from 1948. He now emphasizes a functional form simpler than in his previous study, specifically

$$(1.2) \qquad C = \alpha y_1 + \beta y_2/y_1 + \varepsilon$$

where C is total cost,[2] y_1 is total loaded and empty freight car-miles and y_2 is total loaded freight cars.

Borts proceeded to stratify his data into three size categories and also into three regional categories, eastern, western and southern United States. Using covariance analysis to estimate the cost elasticities, Borts examined the within and between category elasticities. Most striking was the difference across regions as the eastern region exhibited decreasing returns while the southern and western regions were experiencing increasing or constant returns to scale. This difference was attributed to the much higher density of traffic in the eastern region.

---

[2]Cost items joint to passenger and freight services were not included in total costs.

Comparison of the cost elasticities generated within and between size classes supported Borts' contention that a bias towards increasing returns had existed. The between-class elasticities were higher than the within-class elasticities for all regions. Thus estimates of cost elasticities which do not stratify by size will generate cost elasticities that are biased downwards.

Borts concluded by noting that there was still a good chance his estimates were biased as he had estimated a linear total cost function. He points out that if the true cost function is concave from above as is predicted by economic theory, then the cost elasticity will still be biased downward.

The Interstate Commerce Commission (ICC) ruled in 1962 that they would consider 80 percent of railway costs to be variable for regulatory purposes. Griliches (1972) reviews the 'magic number' of 80 percent and finds it considerably below the level indicated by his investigation. Percent variable is defined as the ratio of marginal cost to average cost and is also referred to as the cost elasticity of output. Griliches points out that for both curvilinear and linear cost functions the percent variable varies throughout the length of the cost function. It then follows that the point at which percent variable is calculated can greatly influence the result.

ICC calculations used to establish the 80 percent figure were based on a simple average of the average

cost-output experience of all railways studied. Griliches argues that this is not the correct procedure as it weights equally the experiences of some very large and other quite small railways. Using a weighted average with the proportion of total tons carried as weights Griliches finds that a more correct estimate of percent variable is over 95 percent.

Noting that the ICC studies also incorrectly enter the size of the firm and do not correct for the regression fallacy, as previously noted by Borts (1960), Griliches proceeds to estimate a cost relationship for the period 1957-1960 using cross-sectional data. Railways were divided into two groups due to an observed difference in their cost-output patterns. While the large railways made up just over half the observations, 59 of 97, they accounted for about 95 percent of the total costs reported. Percent variable for the large railways was found to be over 95 percent and not significantly different from 100, thus indicating constant returns to scale. The smaller railways were found to have a percent variable of about 60 percent. Griliches concluded with a review of a number of published studies of railways and noted they in general supported his findings of constant returns to scale.

While the traditional interest in railways has been with regard to returns to scale, more recently the question of excess capacity has been of interest.

Keeler (1974, 1976) investigates both these issues, developing both a short-run and a long-run cost function for railways. Keeler pointed out that many past studies had confused the questions of returns to firm size and returns to density. Proceeding to review Borts (1960) and Griliches (1972), Keeler finds the results from each to be biased due to incorrect specification. He feels that Borts (1960) has a set of "internally inconsistent assumptions." In deriving both short- and long-run cost elasticities from the same data, Borts must be assuming nonoptimal adjustment and optimal adjustment at the same time. Keeler further argues that Borts has incorrectly formed his dependent variable by assuming overhead costs are equal for each ton-mile of passenger and freight service. This leads to a proportional increase in total costs for all roads thus raising the costs of the larger railroads more than the smaller railroads. Keeler points out that this will bias the results toward constant returns to traffic density.

Griliches (1972) is criticized by Keeler in three areas. The first is an incorrectly specified dependent variable which has the same results as those just discussed for Borts. Secondly, Griliches drops the length of track or capacity variable from his estimation based on a regression result where the variable was not significantly different from zero using a 5 percent, two-tailed t-test. Keeler points out the variable is significant

with a 5 percent, one-tailed t-test and feels Griliches has used "a weak reason for throwing out a variable." Finally, Keeler points out that Griliches may not have properly adjusted for heteroscedasticity. Keeler proceeds to demonstrate that, if indeed excess capacity does exist, then Griliches' results will be biased toward constant returns. This follows from the fact that with excess capacity firms will be on short-run cost curves to the left of the long-run cost curve. Estimation of a linear cost function using these observations will then indicate a cost function with a slope greater than the true cost function, therefore biasing the results toward constant returns.

Keeler used cross-sectional data for the United States during the years 1968-1970 to specify a model with two outputs, passenger and freight services, both measured in ton-miles. He assumed that all costs could be charged to one of these two outputs thus allowing the specification of two separate production functions. Using a Cobb-Douglas form for the production functions, it was assumed that the coefficients on the fixed input, track services, were equal for the two production functions. Lagrangean methods were used to construct a short-run cost function for each production function. The identity which required that the total fixed input be equal to the sum of the fixed input allocated to each service, was used to join the cost function into a single short-run cost function.

Keeler, upon testing the degree of homogeneity of his function, found long-run constant returns to firm size were indicated. Differentiating the short-run cost function with respect to track, the fixed input, and setting the result equal to zero, Keeler found an estimate of the optimal capacity for a given traffic level. He substituted this estimate into the short-run cost function to derive a long-run cost envelope. Excess capacity of over 200,000 miles of track was estimated to exist. Increasing returns to traffic density are found to be present over most rail routes in the United States. Keeler concluded that a large saving could be found by abandonment of large amounts of track in the United States. His findings showed that most of this saving was recoverable as only a small portion was sunk costs.

All the above papers point out the joint nature of production in railways but except for the first Borts article no attempt is made to explicitly model this situation. Borts (1952) does model each endogenous input as a function of multiple outputs but his model does not relate the inputs to each other and thus falls short of modeling the full production behavior. Borts (1960) and Griliches (1972) deal only with freight services thus implicitly assuming that a separate production function exists for each of passenger and freight services. Keeler (1974) while introducing two outputs, still assumes they have separate production functions. However, there is a

study which predates all of the above, that explicitly models the joint output nature of railways.

Klein reported such a model in his *Textbook of Econometrics* (1953) which was originally presented by him in a paper for the National Bureau of Economic Research (1947). Modeling a production function with two outputs (passenger and freight services) and three inputs (labour, fuel, and capital services), Klein estimated the following.

$$(1.3) \qquad y_1 y_2^{\delta} = A\, x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} z_1^{\gamma_1} z_2^{\gamma_2}$$

The y's are outputs and the x's are inputs with z_1 being the average length of haul introduced to account for the fixed nature of the track and way and z_2 being the percentage of freight which was the products of mines. This latter variable was introduced as a first step in accounting for the differences in costs of moving various kinds of freight. Klein noted that it would have been desirable to have at least five categories of freight as freight ton-miles are not a homogeneous output.

The output function used by Klein is concave rather than the predicted convex but Klein reported that use of more generally accepted output functions did not change the results of the estimation. Klein used cross-sectional data from 1928 and 1936 to estimate the model. A cost minimization approach was used to estimate the parameters of the production function, thus explicitly

introducing the prices of the inputs into the estimation.[3] Parameter estimates of the production function indicated that increasing returns to scale were present in the railway industry.

There have been a number of reworkings of Klein's model and data since his original estimation. Chipman (1957), Hasenkamp (1976) and Brown *et al.* (1976) have used different estimation procedures and/or different functional forms to examine Klein's original findings. Chipman (1957) used input demand equations from Klein's model to estimate the parameters of the model. Testing for returns to scale Chipman found increasing returns to scale at the one percent level of significance, thus supporting Klein's results.

Hasenkamp (1976) replaced Klein's output function with a constant elasticity of transformation function which allows for convexity of the output set. Two input functions were estimated, first a Cobb-Douglas and secondly a constant elasticity of substitution form. Estimation was undertaken using the system of input demand equations derived by using Shephard's lemma.[4] Hasenkamp imposed the required across-equation restrictions which Chipman

---

[3]Dean (1936) had pointed out that statistical cost estimation should explicitly take account of input prices. Other studies have failed to do this generally assuming that prices were the same for all railways.

[4]See Chapter Two, Section 2.3 for a discussion of Shephard's lemma.

had failed to do but the results did not vary greatly. Increasing returns to scale were again found to be present. Hasenkamp also found that the output structure violated the convexity requirements indicating a concave output structure.

A further reworking of Klein's data was done by Brown *et al.* (1976). As the model used by Brown *et al.* is very similar to that used in the remainder or this thesis a full description is left until the following chapters. Brown *et al.* specify a transcendental logarithmic (translog) functional form for the cost function and then employ Shephard's lemma to derive a system of input demand equations. The system of demand equations plus the cost function is estimated using the translog functional form, allowing Brown *et al.* to test a number of hypotheses which were implicitly implied in earlier specifications. Tests were performed for homogeneity of the production function, separability of inputs and outputs and constant elasticities of substitution. In all cases the hypotheses were rejected indicating the cost function was non-homogeneous, non-separable and did not have constant elasticities of substitution. Earlier estimates are therefore biased due to the functional forms used for the model.

Estimates of returns to scale can be calculated for each firm when homogeneity is not imposed. Brown *et al.* found that all but one of the 66 firms in the sample had significantly increasing returns to scale. Again the

convexity restrictions on the output function were not met, thus indicating the optimal position would be to specialize in one of the two outputs. It is noted by Brown *et al.* that specialization is not possible due to regulation but this does tend to explain the railways' desire to drop passenger services.

1.3 Overview of Thesis

Modeling of multi-product technologies has been greatly advanced recently with the work of Shephard (1970) and Christensen *et al.* (1971). Brown *et al.* (1976) have used these developments in a railway model while Fuss and Waverman (1977) have used them in a study of Bell Canada. Believing that these techniques can be a valuable aid in the discussion of costs in the Canadian railway industry, this thesis develops a model based on the above-mentioned work. The model is estimated using time series data from 1952 to 1976. Ideally, estimation would be based on cross-sectional data from different sections of the railway system over a number of years. Such data, however, are not publicly available.

Chapter Two first presents the duality theory of Shephard (1970) and relates neoclassical production theory and duality theory to the railway industry. Following this is a discussion and proof of Shephard's lemma. The translog cost function is introduced in Chapter Three with

a discussion of its ability to test previously maintained hypotheses. Shephard's lemma is then used to derive a system of cost and revenue share equations from the trans-log cost function. Measures of returns to scale and incremental costs are discussed in the final section of Chapter Three.

A model of the cost relations within Canadian railways is developed in Chapter Four with a review of data sources. Chapter Five discusses the estimation, the tests conducted and presents the results. The final chapter reviews Canadian railway policy in light of the results and its effect on the economy of western Canada.

CHAPTER II

Production theory is the basis for all discussions of pricing and regulation in economics. This chapter examines the received production theory in light of the duality theorems of Shepard (1970). Section one of the chapter discusses the regularity conditions on a standard production correspondence and relates these to the present study of the railway industry. A cost function is defined for the production correspondence and its regularity conditions are discussed in Section Two of the chapter. Shephard's lemma is then proven and discussed in the final section of the chapter.

2.1 Production

Production theory deals with the decisions of the firm given the available technology. The firm marks decisions as to which factors to employ and in what quantities; simultaneously the firm must decide on which output to produce, in what quantities and by what method. It is generally assumed that these decisions are all directed towards the goal of profit maximization, profits being defined as gross revenues minus all opportunity costs.

Given the technology available production theory makes certain predictions as to firms' behavior. These predictions are dependent on both the market structure in which the firm operates and the goal of profit maximization. Under certain technologies and/or market conditions production theory predicts firm behavior which is not optimal for society. Welfare economics calls for the public regulation of these firms with the proper form of regulation depending on the cause of the non-optimal behavior. If regulation is to be employed a good knowledge of the production technology available to the firm is necessary or when it is unavailable it must be estimated. This is done by assuming the observed behavior of firms is consistent with the predictions of production theory. Then working back from the observations the underlying technology can be estimated. Before proceeding to such an estimation for the Canadian railway system, it is necessary to review received production theory and the conditions under which it can be applied.

The typical treatment of production theory assumes a firm producing a single homogeneous output from one or more inputs. Railways do not produce a single output but rather produce a number of non-homogeneous outputs. Outputs can be non-homogeneous due to different qualities or attributes as noted by Spady and Friedlander (1978). Although quality differences may also be present, this study specifies the outputs of railways by attri-

butes. It is assumed that goods exhibiting certain basic attributes will require transportation with different attributes, leading to the four railway outputs defined in this study. The joint nature of production and the non-homogeneity of the output favor the use of a multi-product production function or, as it is also called, a transformation function. Rather than specify directly a transformation function, the discussion that follows will be in terms of a production correspondence as presented in Shephard (1970).

Vectors of inputs and outputs shall be represented by $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_m)$ respectively. Only non-negative input and output vectors will be considered, therefore the sets X and Y can be defined as:

(2.1) $$X = \{x \mid x \geq 0\} = R_+^n$$

(2.2) $$Y = \{y \mid y \geq 0\} = R_+^m$$

The production technologies can then be represented by the production correspondence $T: X \rightarrow Y$ which maps X into Y. This is a point to set mapping; that is, each vector x is mapped into the subset of y vectors that can be produced from x. Given an input vector x, T(x) is the set of production possibilities not all of which need be efficient, efficiency being interpreted as engineering

efficiency; that is, to be called efficient an output vector must be greater than or equal to any other output vector that could be produced from x.

$$y' \text{ is efficient if } y', y'' \varepsilon T(x) \text{ and } y' \geq y'' \text{ for all } y''$$

An inverse correspondence $R:Y \rightarrow X$ can then be defined which maps Y into X. R(y) is the set of input vectors that will produce at least the output vector y. The set R(y) is an isoproduct set, again not all points of which are efficient. Efficiency now requires that the input vector be less than or equal to any other input vector producing output y.

Production theory makes certain assumptions as to the nature of technologies available to the firm. These assumptions can be represented by conditions on the production correspondence. Such a set of conditions are presented by Shephard (1970) and are reproduced below.

T1. $T(0) = \{0\}$.

T2. $T(x)$ is bounded for all $x \varepsilon X$.

T3. $x' \geq x$ implies $T(x') \supset T(x)$.

T4. (a) If $x \geq 0$, $\bar{y} \geq 0$ and $\bar{y} \varepsilon T(\bar{\lambda} x)$ for some scalar $\bar{\lambda} > 0$, then for any scalar $\theta > 0$ there exists a scalar $\lambda^* > 0$ such that $(\theta \bar{y}) \varepsilon T(\lambda^* x)$

or/and

(b) If $x > 0$, or $x \geq 0$ and $\bar{y} \varepsilon T(\bar{\lambda} x)$ for some $\bar{y} > 0$ and $\bar{\lambda} > 0$, then for any $y \varepsilon Y$, $y \varepsilon T(\lambda^* x)$ for some scalar $\lambda^* > 0$.

T5. T is upper semi-continuous on X, implying T(x) closed for all $x \varepsilon X$.

T6. T is quasi-concave on X.

T7. The output sets T(x) of T are convex for all $x \varepsilon X$.

T8. (a) $y \varepsilon T(x)$ implies $\{\theta y | \theta \varepsilon [0,1]\} \subset T(x)$

or/and

(b) $y \varepsilon T(x)$ and $0 \leq y' \leq y$ implies $y' \varepsilon T(x)$.

The first two conditions merely state in mathematical notation the basic underlying problem in all of economics, scarcity. Scarcity is insured as long as no output comes from a zero input vector and only finite amounts of outputs come from finite amounts of inputs.

Free disposal of inputs is allowed for by condition T3. It states that if one or more members of the

input vector are increased then the set of output vectors attainable must be at least as great as it was previously. This insures that no member or members of the input vector can be in such over supply relative to other inputs so as to interfere with the production process. Therefore, the marginal product of any input, while it may reach zero, will never become negative (i.e., allows operation only in the "economic" region of production).

Railway yards can at times become congested as more rolling stock is combined with the fixed track and loading facilities. The preceding condition on the production technology disallows the possibility of this congestion actually reducing the total output. However, the condition does not preclude handling of congestion; a condition reflecting diminishing marginal productivity of fixed factors. Single track mainline may also experience congestion and again marginal product is assumed to be non-negative.

Condition T4 is set down in two parts corresponding to what Shephard calls the "weak and strong attainability of outputs." Part (a) refers to the case where any scalar expansion of an attainable output vector is possible given some scalar expansion of the input vectors used to attain the original output vector. This implies a homogeneous output or there exists some index of outputs which allows output to be expressed as a single value. It then follows that outputs must be produced in

fixed proportions. Therefore any expansion of one output requires the expansion of all outputs by some fixed amount.

The second part of the condition states that a semi-positive input vector which gives a strictly positive output vector can be expanded so as to give all possible output vectors. Output ratios can then be controlled fully by some scalar expansion of some input vector. Desirable outputs can thus be expanded while undesirable ones are controlled or eliminated by a suitable expansion of inputs.

Klein (1953) and others have correctly pointed out that railways do not produce a homogeneous output. Strong attainability of outputs is then required for the modelling of the railway industry. However it should be noted that transportation from A to B always produces in some form the trip from B to A. The existence of the backhaul would indicate that within output classes the assumption of weak attainability might be appropriate. Data are not available to explicitly model the backhaul, therefore strong attainability is assumed both between and within output classes.

The closedness implied by the upper semi-continuity of T insures that the output sets T(x) and the input sets R(y) will be closed. Closure of the sets means that the boundaries are included in the set. This condition allows the definition of the production possibility frontiers and isoproduct curves as the efficient

subsets of the boundaries of the output and input sets, respectively.

The sixth and seventh conditions concern the shape of the production possibility frontiers, the iso-product curves and the transformation function. Quasi-concavity of T insures eventually diminishing marginal product while allowing sections of increasing returns to scale. Shephard (1970) shows that any quasi-concave correspondence will have a convex inverse correspondence, therefore a quasi-concave T implies a convex R. Convexity of R means that there will be diminishing or constant marginal rates of substitution for any input set R(y). Decreasing or constant marginal rates of transformation for any set T(x) are insured by the convexity of the output set.

In considering the railway industry it is vital to have the quasi-concavity of T. Railways experience very large indivisibilities, particularly with respect to track and way. These indivisibilities would indicate a significant region of increasing returns to scale should be allowed for explicitly in the production correspondence.

Shephard (1970) points out that conditions six and seven also allow for "time divisible technologies." This simply means that if two output vectors can be produced by a single input vector, then production of each output vector for a portion of the time interval can obtain a new output vector not otherwise attainable. Thus any convex combination of the attainable output vec-

tors from any single input vector are also attainable from that input vector. Similarly, any convex combination of input vectors, each of which can produce an output vector y, can be used to produce at least the output vector y.

Railway technology is, for the most part, time divisible. The same track, rolling stock and labor can be used over the period of a year to transport a wide variety of goods between a large number of shipping and receiving points. Time divisibility of track and way inputs is essential since most of the railway system in Canada is currently characterized by single-track mainline.

Free disposal of outputs is assumed by condition T8. Parts (a) and (b) indicate the possibilities of assuming weak or strong disposal and correspond to parts T4(a) and T4(b). Weak disposal allows the free disposal of any output only if all members of the output vector are reduced in equal proportion. T8(a) is associated with T4(a) where the ratio of outputs cannot be controlled. Strong disposal allows the free disposal of one or more components of the output vector in varying proportions ranging from 0 to 100 percent of the particular output. The argument why the strong disposal condition, T8(b), is appropriate for a study of railways is exactly parallel to that for strong attainability and need not be repeated here.

2.2 Cost Functions

The production correspondence defines the technological possibilities open to the firm thus constraining the firm's ability to generate profits. Given any input price vector $p = (p_1, p_2, \ldots, p_n)$ and any output price vector $q = (q_1, q_2, \ldots, q_m)$, the firm will adjust its input and output vectors so as to maximize profits. This can be represented by

$$(2.3) \qquad \max_{x,y}\{qy-px \mid y \varepsilon T(x), x \varepsilon R(y)\} p \varepsilon X, q \varepsilon Y$$

Note that p and q are superimposed on the domains of X and Y respectively.

Economists have long realized that for any single output vector y, the profit maximization problem can be expressed as a cost minimization problem for some input price vector p. A cost function corresponding to the production correspondence T can be defined by

$$(2.4) \qquad C(y,p) = \min_{x}\{p \cdot x \mid x \varepsilon R(y)\},\ p \varepsilon X,\ y \varepsilon Y$$

The cost function C(y,p) gives the minimum possible total cost for some output y and input prices p.

Whenever outputs are exogenously determined the cost function is the correct tool of analysis to employ. Railways prior to 1967 faced strong rate regulation, mak-

ing output an exogenous variable. Since 1967, railways have enjoyed somewhat more freedom in setting rates but many rates such as the statutory grain rates are still strictly controlled. Noting that there was a rate freeze during 1973 and 1974 and that the railways have had many commitments to shippers pre-dating the 1967 changes; outputs are assumed to be exogenous throughout the period covered by this study.

When the production correspondence $T:X \rightarrow Y$ has properties T1, ..., T8, the cost function associated with T will have the following properties (Shephard, 1970).

C1. For all $y \geq 0$, $C(y,p)$ is positive and finite for any finite p, $p \neq 0$.

C2. $C(y,\lambda p) = \lambda C(y,p)$ for all $y \varepsilon Y$, $p \varepsilon X$ and $\lambda \varepsilon [0,+\infty]$.

C3. $C(y, p + p') \geq C(y,p) + C(y,p')$ for all $y \varepsilon Y$, $p \varepsilon X$, $p' \varepsilon X$.

C4. $C(y,p') \geq C(y,p)$ for all $y \varepsilon Y$, $p' \geq p \varepsilon X$.

C5. $C(y,p)$ is a concave function of p on X for all $y \varepsilon Y$.

C6. $C(y,p)$ is a continuous function of p on X for all $y \varepsilon Y$.

C7. (a) For any $p \varepsilon X$, $C(\theta y,p) \geq C(y,p)$ for $y \varepsilon Y$ and $\theta \varepsilon [1, +\infty)$

or/and

(b) For any $p \varepsilon X$, $C(y',p) \geq C(y,p)$ for $y' \geq y \varepsilon Y$.

C8. If $p > 0$ and $\{||y^n||\} \longrightarrow +\infty$, then $\lim_{n \to \infty} \text{Inf}\, C(y^n, p) = +\infty$.

C9. For any $p > 0$, $C(y,p)$ is lower semi-continuous in y on Y.

Strictly positive costs for any positive output vector follows from the assumption of scarcity. Finite costs for any finite price vector follows from the fact that all input sets $R(y)$ are bounded, thus cost is the product of two finite vectors.

C2, linear homogeneity of the cost function with respect to prices, is simply the result of optimal input vectors being determined by relative rather than absolute prices. Any scalar expansion of the price vector will leave relative prices and thus the optimal input vector unchanged. Cost will then rise by the same proportion as the prices. This requires that firms do not suffer from any money illusion when making production decisions.

Condition C3 states that given any two or more price vectors, the cost from producing with the optimal

inputs for each price vector will be less than or equal to the total cost from producing with the optimal inputs for any average of the price vectors. This condition is a result of the convexity of the input sets R(y), which imposed constant or diminishing marginal rates of substitution. Following directly from C3 is C4 which says that costs will be monotonicly increasing in prices. An increase in any one or more of the input prices with all other input prices constant must produce a total cost at least as great as that before the price rise.

Concavity of the cost function with respect to prices, C5, is a direct result of C2 and C3. It implies that any convex combination of price vectors will produce a total cost greater than or equal to the total cost of the convex combination of the cost for each price vector. For a case of seasonally fluctuating input prices, this condition shows that the producer would be better off adjusting the vector of inputs to the varying prices rather than trying to stabilize prices at some average level.

It should be noted that the above argument critically depends on the attainment of the optimal input vector for each price vector. Railways have lumpy indivisable inputs that may interfere with a smooth and quick adjustment to any change in price ratios. Therefore, the optimal input vector may not be attainable for some period which means that some convex combination or other averag-

ing of prices may produce a lower total cost.

Economists refer to the situation where a subset of the input vector cannot be readily adjusted to price changes as the short run. C(y,p) is therefore a long-run cost function, where it is assumed all elements of the input vector can be changed to any new optimal level. In utilizing this model, railways are assumed to have overcome any adjustment problems arising from lumpy inputs. This assumption, principally relates to the flow of services from track and way. While it may seem that track and way cannot be adjusted except by total abandonment this is not strictly true. The flow of services from track and way on many rural branch lines have become dependent on the current level of maintenance. This is possibly best exemplified by the current condition of many branch lines in Western Canada. When track is in this condition, the flow of services becomes variable. Supporting this position is Keeler's (1974) finding that only a small portion of way costs are sunk.

The continuity of the cost function for p is a result of there being no holes in the input sets R(y). As long as the technology is time divisible there will be no holes. It has already been noted that railway technology is time divisible and thus this condition will be met.

Condition C7 has two forms depending upon which of conditions T8(a) or T8(b) is assumed. It will be re-

called that T8(a) is the assumption of weak disposability of outputs while T8(b) assumes strong disposability of output. Whichever assumption is made C7 states that the cost function will be monotonically increasing in outputs. Any output vector approaching infinity will have a total cost also approaching infinity for any positive price vector as stated in C8. This is a further scarcity condition. It states that total cost does not reach a limit which would mean that for some finite cost we could get an infinite amount of output.

The final condition on the cost function states that total cost is lower semi-continuous in y. This means that at certain points as output is expanded total cost may make a discrete jump in value. Lumpy input units can be accommodated with this condition as it allows that a new price of rolling stock or yard facilities may be necessary to expand output beyond some point.

As was noted earlier the cost function or the production relation can be used to examine the profit maximizing activities of the firm. What is less well known is that the cost function contains all the necessary information to construct the production structure. Shephard (1970) has proven that given any production structure meeting conditions T1-T8 there is a cost function with properties C1-C9. Conversely, given any cost function with properties C1-C9 there is a production structure meeting conditions T1-T8.

This duality is very convenient for empirical work as it allows the modeling of either the production or cost structures knowing that the properties of both can be tested by testing whichever is modeled. Choosing which to model can then be based on the grounds of data and/or ease of estimation.

2.3 Shephard's Lemma

Shephard's lemma provides a very powerful tool for empirical specification and estimation of any production model. For its implementation we require that the cost function be continuous and differentiable in both input prices and output quantities. Condition C9 on the cost function must be made more restrictive to meet this requirement. The lower semi-continuous assumption made previously must be replaced by a continuous assumption. This will mean that the function no longer allows for discrete jumps in total cost.[5]

Differentiating the cost function with respect to both input prices and outputs, Shephard's lemma states that the following relationships hold.

$$\frac{\partial C(y,p)}{\partial p} = x^* \tag{2.5}$$

$$\frac{\partial C(y,p)}{\partial y} = q \tag{2.6}$$

x^* being the optimal vector of inputs.

---

[5]Condition C9 need not be changed if only the cost function and the input demand equations are used for estimation.

The following proof of Shephard's lemma parallels the proof for the profit function presented by Hasenkamp (1976). First a multiple output production function is defined as

$$(2.7) \qquad F(x,y) = 0$$

Where the inputs are now entered as negative qualities. The following Lagrangian function represents the profit maximization problem.

$$(2.8) \qquad L = qy + px - \lambda[F(x,y)]$$

First order conditions for the maximization of (2.8) are as follows.

$$(2.9) \qquad q - \lambda\frac{\partial F}{\partial y} = 0$$

$$(2.10) \qquad p - \lambda\frac{\partial F}{\partial x} = 0$$

$$(2.11) \qquad F(x,y) = 0$$

Rewriting the cost function in terms of the optimal input vector it becomes:

$$(2.12) \qquad C(y,p) = -px^* \quad \text{with}$$

(2.13) $$x^* = f(p,\bar{y})$$

$\bar{y}$ being the exogenous output vector.

Differentiating C(y,p) with respect to prices and outputs, the following expressions result.

(2.14) $$\frac{\partial C(y,p)}{\partial p} = -x^* - p\frac{\partial x^*}{\partial p}$$

(2.15) $$\frac{\partial C(y,p)}{\partial y} = - p\frac{\partial x^*}{\partial y}$$

Substituting 2.10 into 2.14 and 2.15 yields

(2.16) $$\frac{\partial C(y,p)}{\partial p} = - x^* - \lambda \frac{\partial F}{\partial x} \frac{\partial x^*}{\partial p}$$

(2.17) $$\frac{\partial C(y,p)}{\partial y} = - \lambda \frac{\partial F}{\partial x} \frac{\partial x^*}{\partial y}$$

Now taking the total derivatives of $F(x^*,\bar{y})$ with respect to the vectors of input prices and outputs produces the following:

(2.18) $$\frac{dF(x^*,\bar{y})}{dp} = \frac{\partial x^*}{\partial p} \frac{\partial F}{\partial x} = 0$$

(2.19) $$\frac{dF(x^*,\bar{y})}{dy} = \frac{\partial x^*}{\partial y} \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} = 0$$

Comparing the second expression in 2.16 to 2.18 it is easily seen that 2.16 reduces to 2.5 thus proving the

proposition.

$$\frac{\partial C(y,p)}{\partial p} = -x^* \tag{2.5}$$

Rewriting 2.17 as

$$\begin{aligned} \frac{\partial C(y,p)}{\partial y} &= -\lambda\frac{\partial F}{\partial x}\frac{\partial x^*}{\partial y} - \lambda\frac{\partial F}{\partial y} + \lambda\frac{\partial F}{\partial y} \\ &= -\lambda\left[\frac{\partial F}{\partial x}\frac{\partial x^*}{\partial y} + \frac{\partial F}{\partial y}\right] + \lambda\frac{\partial F}{\partial y} \end{aligned} \tag{2.20}$$

Substituting 2.9 and 2.19 into 2.20 the expression reduces to 2.6 thus proving the second proposition.

$$\frac{\partial C(y,p)}{\partial y} = q \tag{2.6}$$

Empirical work is greatly aided by this, as it is now possible to obtain a system of equations upon specification of a cost function. Estimation becomes more efficient as the number of parameters has not been changed but additional information is available.

CHAPTER THREE

Economic theory is usually developed in terms of implicit functions but before empirical studies can be carried out an explicit functional form must be specified. Since information as to the true functional form is rarely available, it is advantageous to allow for as much flexibility as possible.

All functional form specifications are, in light of a lack of knowledge as to the true form, maintained hypotheses. While this is easily appreciated, what may pass with less notice are any maintained hypotheses which are implied by the functional form specified. The problem of embedded maintained hypotheses has been appreciated for some time, but only recently have functional forms become available which reduce the need to employ as restrictive functional forms. Flexible functional forms now allow testing, by standard statistical methods, of formerly maintained hypotheses such as nonjointness, separability of outputs and inputs, homogeneity and constant elasticities of substitution.

The first section of this chapter discusses nonjointness, separability and homogeneity and their im-

plications for the production technology. Then a translog cost function is specified and certain parameter restrictions are derived. Section Two goes on to derive a system of share equations from the translog cost function and presents tests for the hypotheses discussed in section one. Measures of returns to scale and incremental costs are presented in the final section of the chapter.

3.1 Nonjointness, Separability and Homogeneity

This section examines three properties of production which have often been held as maintained hypotheses in former studies of railway production. As maintained hypotheses, the properties of nonjointness, separability and homogeneity restrict the form the production relation may take. Restrictions embedded in the functional form force the data to conform to a subset of technologies. If the true technology is not a member of this subset then biased estimates will result. What then was the subset of technology that former railway studies assumed? To answer this it is necessary to examine each of the formerly maintained hypotheses.

Jointness in production refers to a situation in which certain factors are inputs to two or more outputs. The following discussion will use jointness to refer to both joint and common cost situations as defined in the regulation literature. Joint cost refers to a situation

in which outputs must be produced in fixed proportions while common costs are defined when output proportions can be varied.

Assuming nonjointness in production implies that the firm can identify fully what proportion of each input is used in the production of each output. Recalling the multi-product production function (2.7) and expanding the input and output vectors produces

$$F(x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_m) = 0 \tag{3.1}$$

Given an assumption of nonjointness (3.1) can be rewritten as follows.

$$\begin{aligned} F_1(x_1, x_2, \ldots, x_n) &= y_1 \\ F_2(x_1, x_2, \ldots, x_n) &= y_2 \\ &\vdots \\ F_m(x_1, x_2, \ldots, x_n) &= y_m \end{aligned} \tag{3.2}$$

It follows from (3.2) that there are m cost functions to correspond to each of the m production functions and that in an empirical study the m equations can each be estimated separately. Given that any single train may carry a wide variety of goods and that many train units will use the same track and switching facilities, it would seem an unreasonable assumption to maintain that railways are characterized by nonjoint production. Thus it would be pre-

ferable to test for nonjointness rather than have it as a maintained hypothesis.

Separability of inputs and outputs has often been assumed in railway studies. This assumption says that there is a homogeneous output which can be expressed as a single value, thus implying an index of outputs exists. When dealing with multiple outputs and joint production this assumes that output substitution possibilities are independent of the level of any input employed. An output index means that the production function can be written as

$$F[h(y),x] = 0 \tag{3.3}$$

where $h(y)$ is the index function. It is further implied from this that some input index $g(x)$ exists such that the production function can be represented as

$$h(y) + g(x) = 0 \tag{3.4}$$

Separability of inputs and outputs has been shown by Lau (1969) to be possible if and only if the cost function is separable in outputs, thus the corresponding cost function is written

$$C(y,p) = C[h(y),p] = 0 \tag{3.5}$$

Implied by this form of the cost function is an assumption

that the relative marginal costs are independent of input prices. Such an assumption is very restrictive if relative input prices change unless all outputs require inputs in the same proportions. This later condition not being very realistic for the railway industry, it would again be preferable to have separability as a testable hypothesis.

Homogeneity of the production function means that for a proportional increase in all inputs, there will be a proportional change in output. This has often been assumed for the single output case, represented as

$$(3.6) \quad Y = f(\lambda x_1, \lambda x_2, \ldots, \lambda x_n) = \lambda^r f(x_1, x_2, \ldots, x_n)$$

where $\lambda \geq 0$ and the function is said to be homogeneous of degree r. A generalization to the multi-product case, which is referred to as almost homogeneity, is suggested by Lau (1969) and takes the form

$$(3.7) \quad F(\lambda^r y_1, \ldots, \lambda^r y_m, \lambda x_1, \ldots, \lambda x_n) = F(y_1, \ldots, y_m, x_1, \ldots, x_n) = 0$$

Almost homogeneity will be referred to as simply homogeneity for the remainder of this thesis.

Brown *et al.* (1976) note that the homogeneity of the production function implies the following for the cost

function.

$$(3.8)\quad C(\lambda^r y,p) = \min_x \sum_{i=1}^{n} p_i \lambda x_i = \lambda \min_x \sum_{i=1}^{n} p_i x_i = \lambda C(y,p)$$

Thus for the production function to be homogeneous of degree r the cost function must be homogeneous of degree $\frac{1}{r}$ in outputs.

It was noted in Chapter Two that condition T6 on the production correspondence allowed for increasing, decreasing and constant returns to scale. An assumption of a homogeneous production function means that only one of the three scale economies can be present. Given that neoclassical production theory predicts in general that production will move from a region of increasing returns through a region of constant returns and on to decreasing returns, the homogeneous specification would bias the results unless all observations fell in the same returns to scale region. Consequently it is more desirable to allow the function to be nonhomogeneous and then test the homogeneity hypothesis. The translog cost function does not completely solve this problem as it is a quadratic function while economic theory would in general predict a cubic total cost function. Estimation may still then be biased if the observations occur over a wide range of a cubic function.

3.2 Translog Cost Function

Maintained hypotheses of the form mentioned above were not imposed purely out of choice in previous production studies. Functional forms which allow these hypotheses to be tested have only recently been developed. Flexible functional forms available include the translog (Christensen *et al.*, 1971), the generalized Leontief (Diewert, 1971), and the generalized Cobb-Douglas (Diewert 1973). The translog form has become the most widely accepted and shall be adopted for this study.

The translog function is a second order Taylor series approximation to an arbitrary cost function. Specifying a cost function of the translog form gives the following expression:

$$(3.9)\quad \ln C = \alpha_0 + \sum_i^n \alpha_i \ln p_i + \sum_j^m \beta_j \ln y_j + \tfrac{1}{2}\sum_i^n \sum_j^n \gamma_{ij} \ln p_i \ln p_j + \tfrac{1}{2}\sum_i^m \sum_j^m \delta_{ij} \ln y_i \ln y_j + \sum_i^n \sum_j^m \rho_{ij} \ln p_i \ln y_j$$

Due to the function being a quadratic there are a large number of parameters to be estimated but this problem is somewhat reduced by the following restrictions.

$$(3.10)\qquad \gamma_{ij} = \gamma_{ji} \qquad i, j = 1, \ldots, n$$

$$(3.11)\qquad \delta_{ij} = \delta_{ji} \qquad i, j = 1, \ldots, m$$

$$(3.12) \qquad \sum_{i}^{n} \alpha_i = 1$$

$$(3.13) \qquad \sum_{i}^{n} \gamma_{ij} = 0 \qquad \text{for all } j = 1, \ldots, n$$

$$(3.14) \qquad \sum_{j}^{n} \gamma_{ij} = 0 \qquad \text{for all } i = 1, \ldots, n$$

$$(3.15) \qquad \sum_{i}^{n} \rho_{ij} = 0 \qquad \text{for all } j = 1, \ldots, m$$

Note that the two adding up constraints on the γ's and the symmetry constraints for the γ's are not independent. That is, the symmetry conditions and any one of the adding up conditions immediately implies the other adding up condition.

The symmetry conditions follow simply from the fact that they are coefficients for the same variables and thus should be equal. Linear homogeneity of the cost function in price, condition C2 from Chapter Two, provides the remaining restrictions as can be seen from the proof below. Only the terms containing prices are reproduced here to save space.

A cost function is linearly homogeneous if and only if a proportional rise in all prices causes an equal proportional rise in costs. Thus if the cost function is to be linearly homogeneous, the following equality must hold when $\lambda > 0$.

$$(3.16) \quad \ln\lambda C = \sum_{i}^{n} \alpha_i \ln \lambda p_i + \tfrac{1}{2} \sum_{i}^{n} \sum_{j}^{n} \gamma_{ij} \ln \lambda p_i \ln \lambda p_j$$
$$+ \sum_{i}^{n} \sum_{j}^{m} \rho_{ij} \ln \lambda p_i \ln y_j + \ldots$$

$$\ln\lambda + \ln C = \ln\lambda\sum_{i}^{n}\alpha_i + \sum_{i}^{n}\alpha_i \ln p_i +$$
$$\tfrac{1}{2}\sum_{i}^{n}\sum_{j}^{n}\gamma_{ij}[\ln p_i \ln p_j + (\ln\lambda)^2 + \ln\lambda p_i + \ln\lambda p_j]$$
$$+ \sum_{i}^{n}\sum_{j}^{m}\rho_{ij}(\ln\lambda\ln y_j + \ln p_i \ln y_j) + \ldots$$

$$\ln\lambda + \ln C = \ln\lambda\sum_{i}^{n}\alpha_i + \sum_{i}^{n}\alpha_i \ln p_i + \tfrac{1}{2}\sum_{i}^{n}\sum_{j}^{n}\gamma_{ij}\ln p_i \ln p_j +$$
$$\tfrac{1}{2}(\ln\lambda)^2\sum_{i}^{n}\sum_{j}^{n}\gamma_{ij} + \tfrac{1}{2}\ln\lambda\sum_{i}^{n}\ln p_i\sum_{j}^{n}\gamma_{ij} +$$
$$\tfrac{1}{2}\ln\lambda\sum_{j}^{n}\ln p_j\sum_{i}^{n}\gamma_{ij} + \sum_{i}^{n}\sum_{j}^{m}\rho_{ij}\ln p_i \ln y_j +$$
$$\ln\lambda\sum_{j}^{m}\ln Y_j\sum_{i}^{n}\rho_{ij} + \ldots$$

$$(3.17) \quad \ln\lambda + \ln C = \ln\lambda\sum_{i}^{n}\alpha_i + \ln C + \tfrac{1}{2}(\ln\lambda)^2\sum_{i}^{n}\sum_{j}^{n}\gamma_{ij} +$$
$$\tfrac{1}{2}\ln\lambda\sum_{i}^{n}\ln p_i\sum_{j}^{n}\gamma_{ij} + \tfrac{1}{2}\ln\lambda\sum_{j}^{n}\ln p_j\sum_{i}^{n}\gamma_{ij} +$$
$$\ln\lambda\sum_{j}^{m}\ln y_j\sum_{i}^{n}\rho_{ij}$$

Upon inspection of (3.17) it can be seen that for the equality to hold equations (3.12) to (3.15) must hold thus providing the restrictions on the cost function.

Estimation is further aided by applying Shephard's lemma as presented in Chapter Two. Shephard's lemma allows the derivation of a system of cost and revenue share equa-

tions from the translog cost function. Differentiating the cost function with respect to the natural logarithms of input prices and outputs the following expressions are derived.

$$(3.18)\qquad \frac{\partial \ln C}{\partial \ln p_i} = \frac{\partial C}{\partial p_i} \cdot \frac{p_i}{C} = \alpha_i + \sum_j^n \gamma_{ij} \ln p_j + \sum_j^m \rho_{ij} \ln y_j \qquad i = 1, \ldots, n$$

$$(3.19)\qquad \frac{\partial \ln C}{\partial \ln y_j} = \frac{\partial C}{\partial y_j} \cdot \frac{y_j}{C} = \beta_j + \sum_i^m \delta_{ij} \ln y_i + \sum_i^n \rho_{ij} \ln p_i \qquad j = 1, \ldots, m$$

Recalling Shephard's lemma and substituting (2.5) and (2.6) into (3.18) and (3.19) respectively,

$$(3.20)\qquad \frac{\partial C}{\partial p_i} \cdot \frac{p_i}{C} = \frac{x_i^* p_i}{C_i} = \frac{\partial \ln C}{\partial \ln p_i} \qquad i = 1, \ldots, n$$

$$(3.21)\qquad \frac{\partial C}{\partial y_j} \cdot \frac{y_j}{C} = \frac{q_j y_j}{C} = \frac{\partial \ln C}{\partial \ln y_j} \qquad j = 1, \ldots, m$$

This procedure produces the following system of n cost share equations and m revenue share equations which are suitable for estimation.

$$(3.22)\qquad \frac{x_i^* p_i}{C} = \alpha_i + \sum_j^n \gamma_{ij} \ln p_j + \sum_j^m \rho_{ij} \ln y_j \qquad i = 1, \ldots, n$$

$$(3.23)\qquad \frac{q_j y_j}{C} = \beta_j + \sum_i^m \delta_{ij} \ln y_i + \sum_i^n \rho_{ij} \ln p_i \qquad j = 1, \ldots, m$$

Including the cost equation there is now a system of $n + m + 1$ equations, which upon the addition of a stochastic specification, is suitable for estimation.

None of the hypotheses discussed in the first section of this chapter are imposed by the translog functional form. Rather tests can be designed to examine these hypotheses. The restrictions for testing the homogeneity of the function in outputs follows directly from equation (3.8) using the same method as the proof for the restrictions for homogeneity in prices. If all outputs are increased by the proportion λ^r, where $\lambda > 0$, and as a result cost increases by the proportion λ, then the cost function is homogeneous of degree $\frac{1}{r}$. For the cost function to be homogeneous of degree $\frac{1}{r}$, the following equality must hold. Again non-essential terms have been dropped.

$$(3.24)\quad \ln\lambda C = \sum_j^m \beta_j \ln\lambda^r y_j + \tfrac{1}{2}\sum_i^m\sum_j^m \delta_{ij}\ln\lambda^r y_i \ln\lambda^r y_j + \sum_i^n\sum_j^m \rho_{ij}\ln p_i \ln\lambda^r y_j + \ldots$$

$$(3.25)\quad \ln\lambda + \ln C = r\ln\lambda\sum_j^m \beta_j + \ln C + \tfrac{1}{2}(r\ln\lambda)^2\sum_i^m\sum_j^m \delta_{ij} + \tfrac{1}{2}r\ln\lambda\sum_j^m \ln y_j \sum_i^m \delta_{ij} + \tfrac{1}{2}r\ln\lambda\sum_i^m \ln y_i \sum_j^m \delta_{ij} + r\ln\lambda\sum_i^n \ln p_i \sum_j^m \rho_{ij}$$

Equality of (3.25) will hold if the following conditions hold.

(3.26) $$\sum_{i}^{m} \delta_{ij} = 0 \qquad j = 1, \ldots, m$$

(3.27) $$\sum_{j}^{m} \delta_{ij} = 0 \qquad i = 1, \ldots, m$$

(3.28) $$\sum_{j}^{m} \rho_{ij} = 0 \qquad i = 1, \ldots, n$$

(3.29) $$\sum_{j}^{m} \beta_{j} = \frac{1}{r}$$

Again only one of the first two conditions is independent due to the symmetry conditions. Thus to test for homogeneity it is only necessary to test the following joint hypotheses.

(3.26) $$\sum_{i}^{m} \delta_{ij} = 0 \qquad j = 1, \ldots, m$$

(3.28) $$\sum_{j}^{m} \rho_{ij} = 0 \qquad i = 1, \ldots, n$$

Should these equalities hold then the cost function will be homogeneous of degree $\sum_{j}^{m} \beta_{j}$ and the production function homogeneous of degree $(\sum_{j}^{m} \beta_{j})^{-1}$. Given the function proved to be homogeneous, a test for constant returns to scale is to test whether $\sum_{j}^{m} \beta_{j} = 1$.

Brown *et al.* (1976) establish a set of conditions for separability of the translog cost function. Recalling that separability implies that the relative marginal costs must be invariant to changes in input prices, the following condition must hold.

$$(3.30) \quad \frac{\partial}{\partial \ln p_h}\left(\frac{\partial \ln C}{\partial \ln y_j} \bigg/ \frac{\partial \ln C}{\partial \ln y_k}\right) = 0 \qquad j,k = 1, \ldots, m \quad h = 1, \ldots, n$$

Using equation (3.19) to substitute into (3.30) the trans-log condition is found to be

$$(3.31) \quad \frac{\partial}{\partial \ln p_h}\left[\frac{\beta_j + \sum_i^m \delta_{ij} \ln y_i + \sum_i^h \rho_{ij} \ln p_i}{\beta_k + \sum_i^m \delta_{ik} \ln y_i + \sum_i^n \rho_{ik} \ln p_i}\right] = 0$$

A sufficient condition for (3.31) to hold is that

$$(3.32) \quad \rho_{ij} = 0 \qquad i = 1, \ldots, n \quad j = 1, \ldots, m$$

Brown *et al.* use this set of restrictions to test for separability noting that only m(n-1) of the ρ_{ij}'s are independent due to condition (3.15) being imposed. If some $\rho_{ij} \neq 0$ then separability may still hold but this would require that all the ratios of cost elasticities be invariant to the levels of outputs and prices. Brown *et al.* choose not to examine this possibility due to its very restrictive nature.

Testing for nonjointness is more difficult in the translog form as it can only be done with an approximation test as discussed by Denny and Fuss (1977). This test requires that the data be scaled so as $p_i = y_j = 1$ at the point of approximation. Given this condition Denny and Pinto (1978) have shown that the necessary re-

strictions to test for nonjointness are

$$(3.33) \qquad \delta_{ij} = -\beta_i\beta_j \qquad i,j = 1, \ldots, m \quad i \neq j$$

3.3 Returns to Scale and Incremental Cost

Returns to scale have traditionally been measured by examining the output cost elasticity. This is a satisfactory measure when there is a homogeneous output but does not easily generalize for the case of multiple outputs. Fuss and Waverman (1977) discuss this question and point out that the correct value to examine is

> the change in total cost resulting from differential changes in the levels of the m outputs.[6]

For the translog cost function the expression for the above mentioned change is

$$(3.34) \qquad d\ln C = \sum_{i}^{m} \frac{\partial \ln C}{\partial \ln y_i} \, d\ln y_i$$

Further specification is required for this expression to be interpreted as a measure of returns to scale. The specification suggested by Fuss and Waverman (1977) is that

---

[6]Melvyn Fuss and Leonard Waverman, "Multi-product Multi-input Cost Functions for a Regulated Utility: The Case of Telecommunications in Canada." (Paper presented at the National Bureau of Economic Research Conference on Public Regulation, Washington, December, 1977), p. 10.

$$(3.35) \qquad d\ln y_i = \lambda \qquad\qquad i = 1, \ldots, m$$

λ is a constant

This states that the measurement is to be performed when a proportional increase in all outputs is considered. Substituting (3.35) into (3.34) the expression for returns to scale is

$$(3.36) \qquad \frac{d\ln C}{\lambda} = \sum_{i}^{m} \frac{\partial \ln C}{\partial \ln y_i}$$

The returns to scale are then indicated by (3.36) being greater than, less than or equal to one. If $\frac{d\ln C}{\lambda} < 1$, then there are increasing returns to scale, with constant and decreasing returns being indicated by $\frac{d\ln C}{\lambda} = 1$ and $\frac{d\ln C}{\lambda} > 1$, respectively.

Fuss and Waverman (1977) go on to demonstrate that there does not exist a measure of returns to individual outputs that is consistent with the measure of overall returns as defined by (3.36). All attempts to develop a measure of returns for individual outputs lead to the possibility of having increasing returns in all single outputs but decreasing returns overall. This inconsistency cannot be overcome and actually becomes more probable when the number of outputs to be considered increases.

While the cost elasticities of individual outputs are not generally useful in examining returns to scale, due to the above, Fuss and Waverman (1977) demonstrate

that they can be used to develop a measure of incremental cost. Incremental cost is the change in total cost due to an increase in one output holding all other outputs constant in a joint product technology. This differs from marginal cost as truly joint costs cannot be causally related to any single output. The translog cost function provides individual cost elasticities of the following form:

$$\frac{\partial \ln C}{\partial \ln y_i} = \beta_i + \sum_j^m \delta_{ij} \ln y_j + \sum_j^n \rho_{ij} \ln p_j \qquad i = 1, \ldots, m \tag{3.37}$$

Rewriting the expression on the left of the equality and rearranging terms, an expression for $\frac{\partial C}{\partial y_i}$ results.

$$\frac{\partial \ln C}{\partial \ln y_i} = \frac{\partial C}{\partial y_i} \frac{y_i}{C} \tag{3.38}$$

$$\frac{\partial C}{\partial y_i} = \frac{C}{y_i} \frac{\partial \ln C}{\partial \ln y_i} \qquad i = 1, \ldots, m \tag{3.39}$$

Evaluating (3.39) with the levels of input prices and other outputs held constant, an expression for the incremental cost curve (ICC) is obtained. Fuss and Waverman (1977) present the incremental cost curves for the translog cost function.

$$ICC(y_i) = \frac{C}{y_i} \left[\beta_i + \delta_{ii} \ln y_i + \sum_{j \neq i}^{m-1} \delta_{ij} \ln \bar{y}_j + \sum_j^n \rho_{ij} \ln \bar{p}_j \right] \qquad i = 1, \ldots, m \tag{3.40}$$

Where $\bar{y}_j$ and $\bar{p}_j$ are preassigned values of outputs and input prices; usually at an average level.

The translog functional form, upon the addition of a stochastic disturbance, allows a more general modeling than was previously possible. Chapter Four proceeds to use the specification outlined in this chapter to develop a framework for analysing the cost relationships in Canadian railways.

CHAPTER FOUR

Railways are characterized by a multiple input, multiple output production process. It is assumed that the production process can be represented by a single production function, rather than a system of production functions as specified in Borts (1952). Most previous studies have concentrated on the input set of the railway industry while assuming one or possibly two outputs. Klein (1953) pointed out that five or more categories of freight services would be required for a complete study of the railway industry. The present study extends the examination of the output set to include four outputs. Specifying four outputs was only possible with a simplified, two factor, input set due to data limitations. While it would have been preferable to estimate the model with cross-sectional data from different track sections on the various railways, the only public data were for the country as a whole on a time series basis.

4.1 A Cost Function for Canadian Railways

The Canadian railways are assumed to be characterized by a production function using two inputs and pro-

ducing four outputs. Labor services (L), measured as millions of man-hours, and capital services (K), measured as millions of dollars of expenditure on materials for operations and maintenance plus depreciation, are taken to be the inputs of the railways. Outputs are measured by ton-miles of service for the four outputs specified. The output categories are defined by the type of cargo being transported with the four cargo types being

1. statutory grains (S),
2. agricultural products (A),
3. products of forests and mines (F), and
4. manufactured goods (M).

Statutory grains are defined in the National Transportation Act while categories two, three and four correspond to the classifications used by the Canadian Transport Commission prior to 1969.

Specification of a cost function requires the prices of the inputs and the total cost of production (C) in addition to the output quantities. The price of labor was taken to be the average hourly compensation paid to railway employees in each year. There was no readily available price for capital but on examination it was found that expenditures on materials made up the greatest portion of this variable. This being the case, the general wholesale price index was used as a proxy for the price of capital. Total cost was measured by the total expenditures on labor and materials plus the depreciation on track and

rolling stock. In addition to the above variables, the revenue from each output is necessary if the system of share equations are to be used.

Hicks' neutral technical change is assumed for railway production.[7] This assumption implies that the level of the cost function changes by some fixed proportion over time. Hicks' neutral technical change leaves the substitution possibilities within and between the input and output sets unchanged. To allow for the shifting of the cost function due to technical change, time is introduced by a vector of integers 1-25, corresponding to the years 1952-1976.

With this final variable, the cost function can be written as

$$(4.1)\quad \ln C = \alpha_0 + tT + \sum_i^2 \alpha_i \ln p_i + \sum_j^4 \beta_j \ln y_j + \tfrac{1}{2}\sum_i^2\sum_k^2 \gamma_{ik} \ln p_i \ln p_k + \tfrac{1}{2}\sum_j^4\sum_h^4 \delta_{jh} \ln y_j \ln y_h + \sum_i^2\sum_j^4 \rho_{ij} \ln p_i \ln y_j$$

$$i,k = L,\ K$$

$$j,h = S,\ A,\ F,\ M$$

where p and y are input prices and outputs and where α_0, t, α_i, β_j, γ_{ik}, δ_{jh}, ρ_{ij} are parameters to be estimated. Using Shephard's Lemma, two cost share (CS) and four revenue share (RS) equations can also be defined.

$$(4.2)\qquad CS_i = \alpha_i + \sum_k^2 \gamma_{ik} \ln p_k + \sum_j^4 \rho_{ij} \ln y_j$$

---

[7]See Oum (1976) for an example of a translog function with nonneutral technical change.

$$(4.3) \qquad RS_j = \beta_j + \sum_h^4 \delta_{jh} \ln y_h + \sum_i^2 \rho_{ij} \ln p_i$$

$$i,k = L,\ K$$

$$j,h = S,\ A,\ F,\ M$$

Imposing the restrictions outlined in Chapter Three and the necessary across-equation restrictions will complete the specification of the cost relations. The symmetry conditions and the homogeneity in prices provide the following restrictions.

$$\gamma_{ij} = \gamma_{ji} \qquad i,j = L,\ K$$

$$\delta_{ij} = \delta_{ji} \qquad i,j = S,\ A,\ F,\ M$$

$$\sum_i^2 \alpha_i = 1 \qquad i = L,\ K$$

$$\sum_i^2 \gamma_{ij} = 0 \qquad i = L,\ K$$

$$\sum_i^2 \rho_{ij} = 0 \qquad i = L,\ K$$

When a coefficient appears in more than one equation in the system, a restriction making its value the same in each equation should be imposed.

4.2 Data Specification

The cost function and the share equations were estimated using data from publications of Statistics Canada and the Canadian Transport Commission.[8] Cost and input

---

[8]Formerly the Dominion Bureau of Statistics and the Board of Transport Commissioners of Canada.

data were taken from the six-part Statistics Canada publication, *Railway Transport*. The Canadian Transport Commission's publication, *Waybill Analysis Carload All-Rail Traffic*, was the source of the output data.

Labor inputs were reported as total hours of employment for all employees. The total compensation paid to this labor was also reported as well as the average hourly compensation which was a simple average of the above two values. As these data were in current dollar values, they were adjusted by the general wholesale price index where 1964 = 100. This index was chosen so that the labor expenditures would be comparable to the capital expenditures.

Total expenditures on operations and maintenance plus depreciation were reported in Part II of *Railway Transport*. The operations and maintenance expenditures included labor costs which were netted out to get the expenditure to be included in capital costs. As the price of materials making up the above expenditure was taken to be the general wholesale price index, this index was used to adjust the current dollar values of these expenditures.

Indexing the depreciation costs presented a further problem. Depreciation is calculated on historical values and charged in the current period. This means that the depreciation charge for any one year is made up of dollar values from a number of past years, typically about 20 for the railways. No breakdown of depreciation by year

of purchase was available, so a simple average of the general wholesale price index was used to adjust the depreciation costs of a given year to their 1964 equivalent. This is equivalent to assuming that the depreciation in each year is equally distributed over the average life of all depreciable assets. The average life of depreciable assets was taken to be 20 years prior to 1968 and 18 years subsequently. These figures were based upon an examination of the average depreciation rate charged by the Canadian Pacific Railway. Prior to 1968, the average depreciation rate was approximately 5 percent, indicating an average life of 20 years. From 1968 onwards, the rate was approximately 5.5 percent indicating an 18-year life for the assets. It was then assumed that these depreciation rates were characteristic of all Canadian railways.[9]

Output data were taken from the *Waybill Analysis Carload All-Rail Traffic* published by the Canadian Transport Commission. This is a sample of car load freight moving within Canada which excludes trans-border shipments, less than car load lots and rail-water-rail shipments. Data on revenue received, tons carried, ton-miles and revenue per ton-mile are presented.[10]

As these data are from a sample whereas the cost

---

[9]The general wholesale price index for 1934 was not available. A straight line averaging of 1933 and 1935 was used.

[10]In excluding the traffic types mentioned in the text, the sample excludes about 30 percent of all rail traffic.

and input data are for the total traffic, some scaling of the data was necessary. Scaling was accomplished by raising the revenue and ton-mile values by the proportion of total tons carried to the sample tons carried for each freight class. Data on total tons carried by freight class are available from Statistics Canada. This scaling assumes that freight movements excluded from the sampling are distributed in proportion to the sampled freight for each class. Revenue data was adjusted by the general wholesale price index to the same base year as the cost data.

4.3 Input and Output Substitution

Substitution among factors or products by firms and consumers in the face of relative price changes is a fundamental concern of economics. When faced with changing market conditions, the firm must decide how to react by substituting one input for another, one output for another or both. Input substitution has been dealt with on many occasions but as few previous studies allow for multiple outputs, the possibilities for output substitution have been less well examined in empirical studies. Firms' reaction to price changes are of primary interest and to examine these the elasticities of substitution and demand and supply elasticities are examined.

Allen (1938) defined a partial elasticity of substitution (APES) between two inputs. This is a measure of the percentage change in factor proportions to a per-

centage change in their price ratio. Allen originally defined this measure in regard to a production function

$$(4.4) \qquad f(x_1, \ldots, x_n, y_{n+1}, \ldots, y_{n+m}) = 0$$

where x and y are inputs and outputs respectively. Writing the partial derivatives of this production function as f_i and the second order partial derivatives as f_{ij}, Allen's partial elasticity of substitution can be written as

$$(4.5) \quad \sigma = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}{x_i x_j} \frac{F_{ij}}{F}$$

$$i,j = 1, \ldots, n$$
$$i \neq j$$

$$\text{where } F = \begin{vmatrix} 0 & f_1 & f_2 & f_3 & \ldots & f_n \\ f_1 & f_{11} & f_{12} & & & \cdot \\ f_2 & f_{21} & f_{22} & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \\ \cdot & & & & & \\ f_n & \cdot & \cdot & \cdot & & f_{nn} \end{vmatrix}$$

and F_{ij} is the co-factor of the term f_{ij} in F.

Uzawa (1962) has shown that this can be written in terms of a cost function, as

$$(4.6) \qquad \sigma_{ij} = \frac{C \dfrac{\partial^2 C}{\partial p_i \partial p_j}}{\dfrac{\partial C}{\partial p_i} \dfrac{\partial C}{\partial p_j}} \qquad i,j = 1, \ldots, n \quad i \neq j$$

Bernt and Wood (1975) have taken Uzawa's result and found that for the translog cost function the elasticity of substitution can be written as

$$(4.7) \qquad \sigma_{ii} = \frac{\gamma_{ii} + CS_i^2 - CS_i}{CS_i^2} \qquad i = 1, \ldots, n$$

$$(4.8) \qquad \sigma_{ij} = \frac{\gamma_{ij} + CS_i CS_j}{CS_i CS_j} \qquad i,j = 1, \ldots, n \quad i \neq j$$

It should be noted that σ_{ii} does not in itself have any interpretation but is calculated because it is useful in determining the own price elasticity of demand for the factor.

Allen (1938) demonstrates the relationship between his partial elasticity of substitution and the price elasticity of demand. The price elasticity of demand being defined as

$$(4.9) \qquad E_{ij} = \frac{\partial \ln x_i}{\partial \ln p_j} \qquad i,j = 1, \ldots, n$$

when output and other input levels are held constant. The elasticity of demand is a simple function of the APES and the cost shares.

$$E_{ij} = CS_j\sigma_{ij} \qquad i,j = 1, \ldots, n \tag{4.10}$$

Bernt and Wood (1975) point out that whereas $\sigma_{ij} = \sigma_{ji}$, it is not generally the case that $E_{ij} = E_{ji}$.

Allen (1938) states that it is possible to develop a parallel measure for the substitution between outputs. The partial elasticity would now measure the percentage change in output ratios to a percentage change in their price ratios. Referring to this elasticity as the Allen partial elasticity of transformation (APET), it can be defined as the following using (4.4).

$$\tau_{ij} = \frac{y_{n+1}f_{n+1} + y_{n+2}f_{n+2} + \ldots + y_{n+m}f_{n+m}}{y_iy_j}\frac{D_{ij}}{D} \tag{4.11}$$

$$i,j = 1, \ldots, m$$
$$i \neq j$$

where $D = \begin{vmatrix} 0 & f_{n+1} & f_{n+2} & f_{n+3} & \cdots & f_{n+m} \\ f_{n+1} & f_{n+1,n+1} & f_{n+1,n+2} & & & \\ f_{n+2} & f_{n+2,n+1} & f_{n+2,n+2} & & & \\ \vdots & & & & & \\ f_{n+m} & & \cdots & & & f_{n+m,n+m} \end{vmatrix}$

and D_{ij} is the co-factor of the term $f_{n+i,n+j}$ in D.

It can be shown (see Appendix A) that this elasticity can be written as

$$(4.12) \qquad \tau_{ij} = \frac{R \dfrac{\partial y_i}{\partial q_j}}{y_i y_j} \qquad i,j = 1, \ldots, n \quad i \neq j$$

where R is total revenue. This can also be expressed in terms of a cost function by

$$(4.13) \qquad \tau_{ij} = \frac{R}{y_i y_j} \cdot \frac{1}{\dfrac{\partial^2 C}{\partial y_i \partial y_j}} \qquad i,j = 1, \ldots, n \quad i \neq j$$

Solving (4.13) in the case of the translog cost function results in

$$(4.14) \qquad \tau_{ij} = \frac{R}{C\delta_{ij}} \qquad i,j = 1, \ldots, n \quad i \neq j$$

Price elasticities of supply for the outputs can be shown to be a function of the APET (see Appendix A). With the price elasticity of supply being defined as

$$(4.15) \qquad ES_{ij} = \frac{\partial \ln y_i}{\partial \ln q_j} \qquad i,j = 1, \ldots, n$$

The supply elasticities are fully determined by the revenue shares and the APET.

$$(4.16) \qquad ES_{ij} = K_j \tau_{ij} \qquad i,j = 1, \ldots, n$$

Parallel to the substitution case $\tau_{ij} = \tau_{ji}$ but in general $ES_{ij} \neq ES_{ji}$.

Own price elasticities of demand and the σ_{ii}'s are expected to be negative in sign while the sign of E_{ij} and σ_{ij}, $i \neq j$ will depend on the factors being net substitutes or complements -- substitutes giving a positive sign; complements a negative sign. The expected signs are the opposite for the output set. ES_{ii} and τ_{ii} are expected to be positive while ES_{ij} and τ_{ij}, $i \neq j$ are expected to be negative for substitutes and positive for complements.

CHAPTER FIVE

This chapter presents the empirical results from estimation of the cost relations outlined in Chapter Four. Section 5.1 discusses the estimation procedure used, presents coefficient estimates and tests for homogeneity and separability of the production function. Measures of returns to scale and incremental costs are presented in the second section of the chapter. The final section, 5.3, presents estimates of the Allen partial elasticities of substitution and transformation and factor demand and output supply elasticities with a discussion of their interpretation.

5.1 Estimation of the Cost Relations

Estimation of the cost and revenue share equations and the cost function outlined in Chapter Four followed the procedure used by Christensen and Greene (1976). First a disturbance term was added to each of the structural equations. The error term is added to capture three effects on the equations: (1) omitted variables, assumed to each have a small effect but in total is worth noting, (2) measurement error in the dependent variable, (3) pure

randomness in economic behavior. It was assumed that the disturbances were jointly normally distributed and additive. Additive disturbances assume the errors are a linear extension of the specified equations while the normal distribution is required for the test statistics to have the proper distributions.

Zellner's seemingly unrelated regressions technique would seem to be the proper procedure to use for the system of seven equations as the disturbance terms for each year are likely to be affected by the same economic conditions. However, Christensen and Greene (1976) point out that the system of equations as it is currently specified will have a singular estimated disturbance covariance matrix. This result occurs because the errors on the cost share equations sum to zero since the cost shares must sum to one by definition. To avoid the singular covariance matrix one of the cost share equations must be dropped from the estimation. The estimation results will in general depend on the equation dropped, however, Christensen and Greene (1976) note that if a maximum likelihood estimator is used, the parameter estimates are not affected by the equation dropped from the system. The method employed here involves iterating on the Zellner technique which generates maximum likelihood estimates upon covergence. The system was estimated using the iterative Zellner technique with the capital cost share equation dropped from

the system. It should be noted that the estimates will be slightly biased due to the exclusion of passenger service from the outputs. The total costs include expenditures on passenger service but as passenger service makes up a small proportion of the total railway costs and revenues this bias should be small. Observed costs will be above the true cost function for each output level but as passenger service has been a declining proportion of railway output, this upward bias diminishes over the study period. This will tend to bias downwards the estimated overall returns to scale and the calculated incremental costs, particularly in the early years as the slope with respect to each output will be lower and the average costs higher at each level of output. Incremental costs will appear to rise more quickly or fall more slowly with output than in actual fact due to the flattening of the observed cost function. Overall returns to scale will be affected by both the decreased slope and the increased average cost. This is because the overall returns to scale measure is the sum of the individual cost elasticities which vary directly with the slope and inversely with average cost.

The Allen partial elasticities of transformation will also be biased downwards. These elasticities depend on the inverse of the coefficient on the cross product terms in outputs [see equation (4.14)]. With the observed cost function lying above the true cost function but approaching it as passenger output declines, the slope

of the observed cost function will be rising at a rate faster than the true cost function. This means the estimated coefficients will be larger than the true coefficients and therefore the estimated APET will be smaller than the true elasticities.

Estimated values of the coefficients of the cost function are reported in Table 5-1 along with their standard errors and asymptotic t-statistics. Asymptotic t-statistics are reported since the test statistic will have a t-distribution only in the limit as the number of observations goes to infinity. An R^2 was calculated for each of the six equations and these are reported in Table 5-2. The relatively low R^2 for the statutory grain revenue share equation is due to a very unstable revenue share which follows from highly variable grain shipments due to fluctuations in output and world demand. Since Canada only supplies about five percent of the world grain market, its export demand curve is very elastic causing wide variations in exports with world supply conditions. Canada's ability to meet export demand also varies widely depending upon the recent harvests and exports.

The coefficient for Hicks' neutral technical change, t, indicates an exogenous drop of slightly more than 1 million dollars per year in the level of total cost. This is equivalent to a fall of about .04 percent per year. Tests were performed for homogeneity of production and separability of inputs and outputs. The test statistics

Table 5-1

ESTIMATED TRANSLOG COST FUNCTION

Regression Coefficient	Estimated Value	Standard Error	t-Statistic
α_0	7.9582	.03694	215.740*
α_L	.59843	.00325	184.120*
α_K	.40157	.00325	123.560*
β_S	.08955	.00169	52.860*
β_A	.05489	.00101	54.250*
β_F	.23088	.00456	50.610*
β_M	.43323	.00879	49.310*
γ_{LL}	.09727	.01277	7.615*
γ_{LK}	-.09727	.01277	-7.615*
δ_{SS}	.07053	.00571	12.350*
δ_{SA}	-.00887	.00302	-2.935*
δ_{SF}	-.02338	.00775	-3.018*
δ_{SM}	-.01347	.00837	-1.610**
δ_{AA}	.04213	.00537	7.845*
δ_{AF}	.00169	.00534	.315
δ_{AM}	-.02370	.00572	-4.142*
δ_{FF}	.19517	.02124	9.190*
δ_{FM}	-.10411	.02276	-4.575*
δ_{MM}	.22110	.02777	7.961*
ρ_{LS}	.02268	.00540	4.202*
ρ_{LA}	.01368	.00350	3.906*
ρ_{LF}	-.03922	.01116	-3.515*
ρ_{LM}	.00099	.01243	.795
ρ_{KS}	-.02268	.00540	-4.202*
ρ_{KA}	-.01368	.00350	-3.906*
ρ_{KF}	.03922	.01116	3.515*
ρ_{KM}	-.00099	.01243	-.795
t	-.07001	.00224	-31.300*

Number of observations = 25.

*Significant at the one percent level of confidence, two-tailed test.

**Significant at the six percent level of confidence, one-tailed test.

Table 5-2

GOODNESS OF FIT STATISTICS

Equation	R^2
Cost function	.999
Labor cost share	.999
Statutory grain revenue share	.746
Agriculture revenue share	.997
Forests and mines revenue share	.993
Manufactured revenue share	.992

are from an F-distribution with the values and degrees of freedom reported in Table 5-3. Both of the test statistics are significant at the one percent level, leading to a rejection of the null hypothesis of homogeneity and separability. Brown *et al.* (1976) also reject the hypothesis of homogeneity and separability and as noted by them, this leads to a rejection of functional forms used in earlier studies. Rejecting homogeneity allows the returns to scale to vary throughout the length of the cost function. Non-separable production indicates that there is not a homogeneous output, requiring the explicit modeling of multiple outputs and different input-output substitution possibilities for each output.

5.2 Returns to Scale and Incremental Costs

Overall returns to scale were estimated for each of the 25 years in the sample using the overall cost elasticity given by (3.36). The estimated values and their 95 percent confidence intervals are reported in Table 5-4. While the point estimates are all less than one, indicating increasing returns to scale, the confidence intervals for 1969, 1971-1976 include regions of increasing, constant and decreasing returns to scale. Even when the small bias mentioned earlier is allowed for, increasing returns are indicated for the early years of the sample with a trend toward constant and perhaps decreasing returns.

Table 5-3

TESTS FOR HOMOGENEITY AND SEPARABILITY

Null Hypotheses	F-Statistic	DF1	DF2
Homogeneous production function	22.318*	5	128
Separable production function	19.120*	4	128

*Significant at the one percent level of confidence.

Table 5-4

MEASURE OF RETURNS TO SCALE

Year	Returns to Scale	95 Percent Confidence Interval	
1952	.729	(.893	.564)
1953	.711	(.907	.514)
1954	.698	(.911	.486)
1955	.731	(.892	.570)
1956	.760	(.882	.639)
1957	.762	(.872	.653)
1958	.743	(.883	.604)
1959	.746	(.885	.608)
1960	.733	(.892	.574)
1961	.741	(.885	.596)
1962	.757	(.879	.635)
1963	.779	(.861	.697)
1964	.809	(.840	.777)
1965	.824	(.901	.748)
1966	.829	(.894	.763)
1967	.840	(.939	.740)
1968	.844	(.972	.715)
1969	.860	(1.022	.698)
1970	.861	(.980	.741)
1971	.897	(1.073	.720)
1972	.893	(1.065	.722)
1973	.924	(1.143	.705)
1974	.938	(1.183	.693)
1975	.941	(1.188	.693)
1976	.931	(1.161	.700)

Keeler (1976) warned that often discussions of returns to scale confused returns to density and returns to plant size. The returns measure employed here is a function of output levels therefore measuring returns to density. Keeler (1976) also found increasing returns to density. This does not mean that all sections of railway in Canada are experiencing increasing returns to density. Some sections of track, particularly mainline track, may be in a decreasing returns situation while other sections may have a large range of output levels over which increasing returns still persist. A study using cross-sectional data would allow the estimation of returns to density for different sections of line possibly indicating where variations in rates across regions might be justified.

Incremental cost curves have been calculated for each of the outputs and presented in Figures 5-1 to 5-4. The incremental costs for statutory grain and agricultural products are constant or slightly increasing while the incremental costs for manufactured goods clearly decline. This result is possibly due to a high proportion of fixed costs being attributed to manufactured output. With the large increase in manufactured output over the period of the study, incremental costs would fall and the large increase in forest and mining output may have meant this category provided revenues for a larger portion of fixed costs.

Two incremental cost curves are suggested by

Figure 5-1

INCREMENTAL COSTS
STATUTORY GRAIN

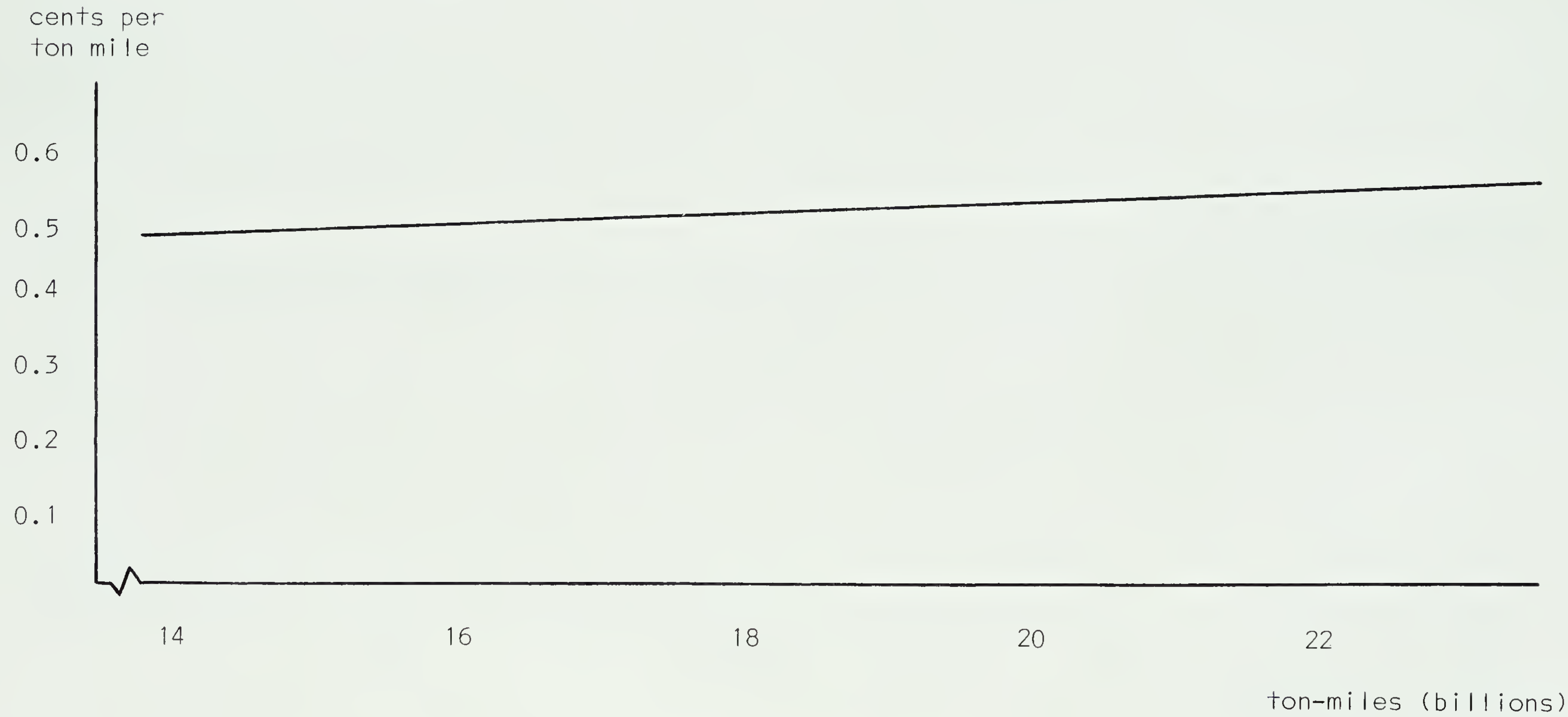


Figure 5-2

INCREMENTAL COSTS
AGRICULTURAL PRODUCTS

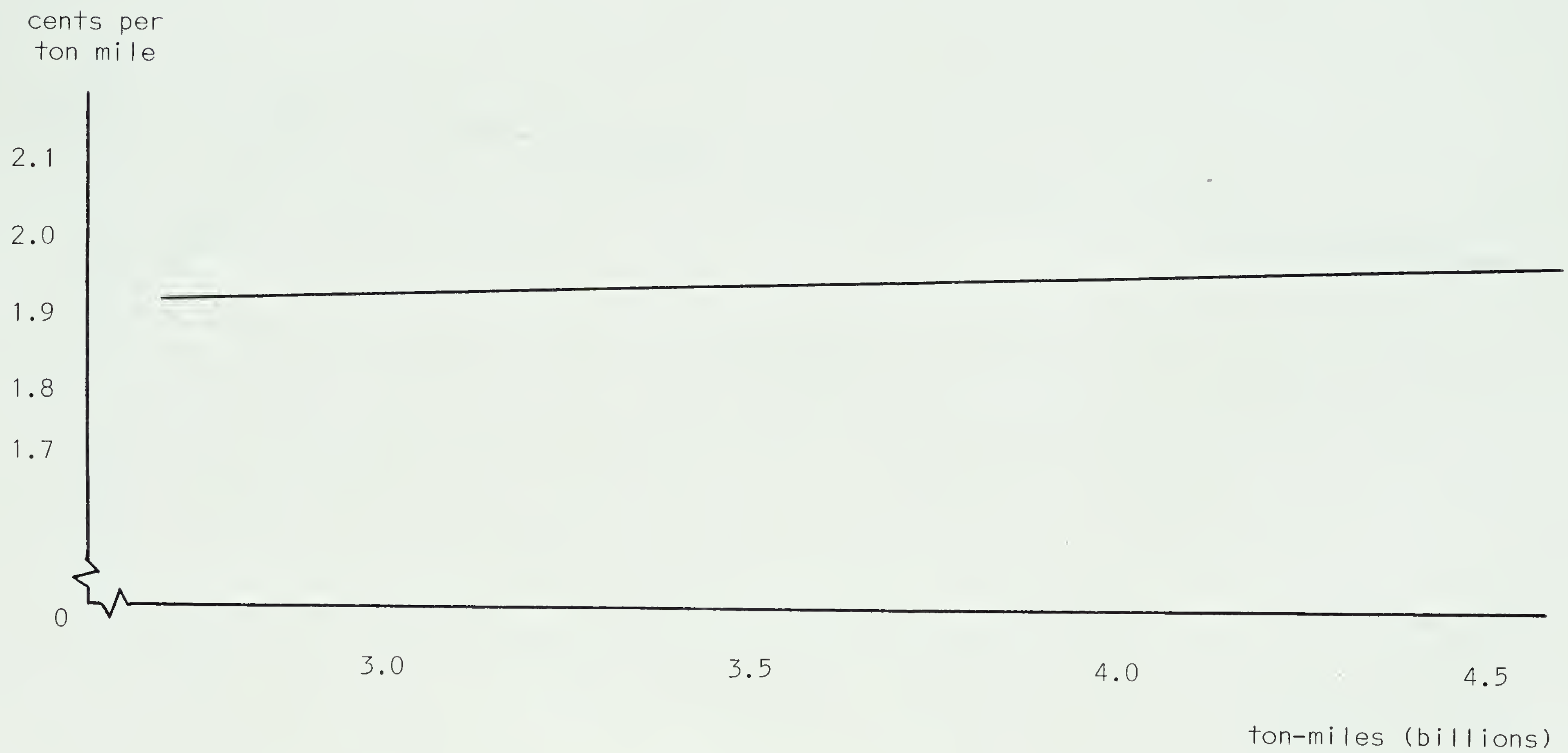


Figure 5-3

INCREMENTAL COSTS
FOREST AND MINING PRODUCTS

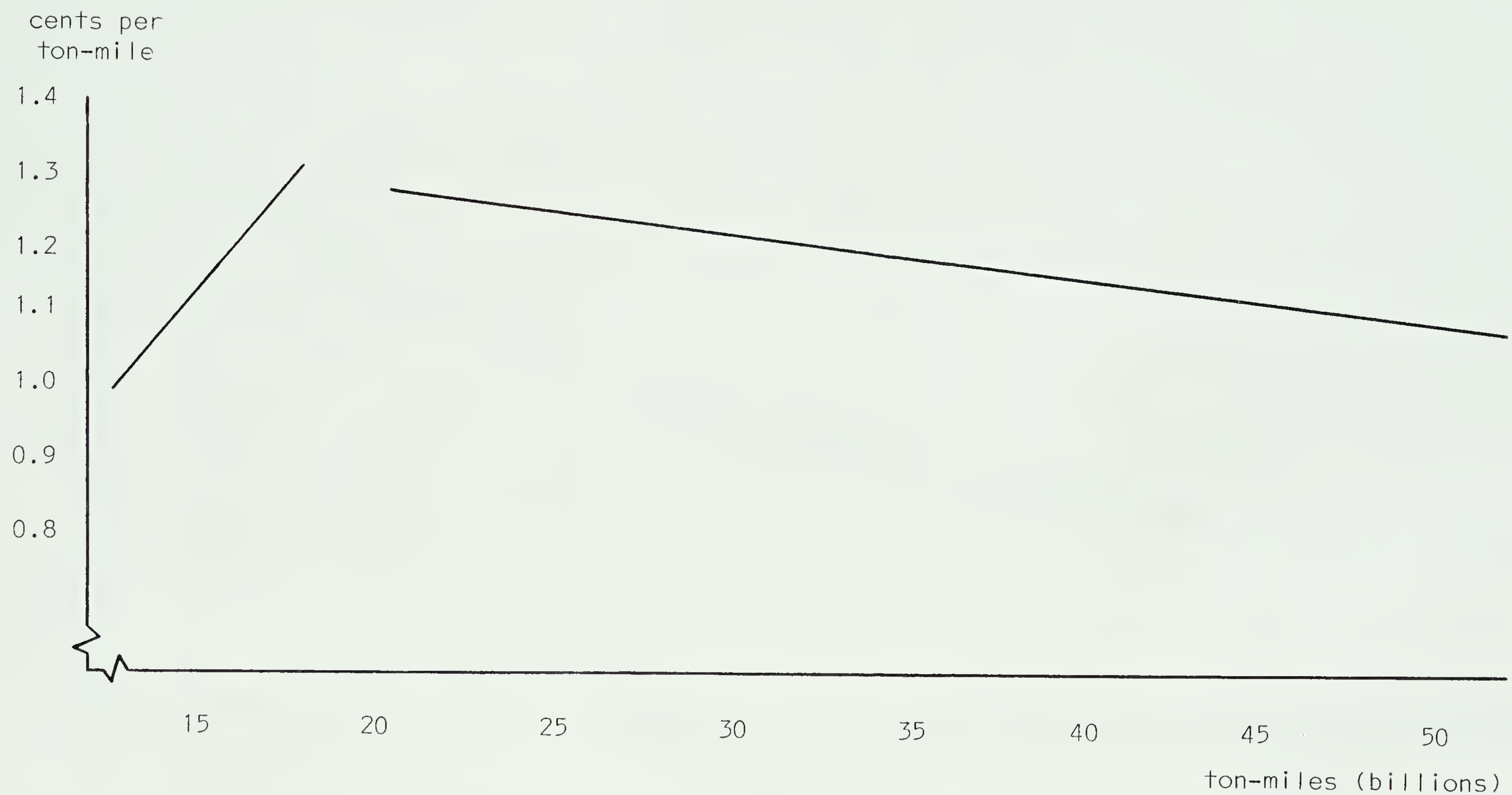
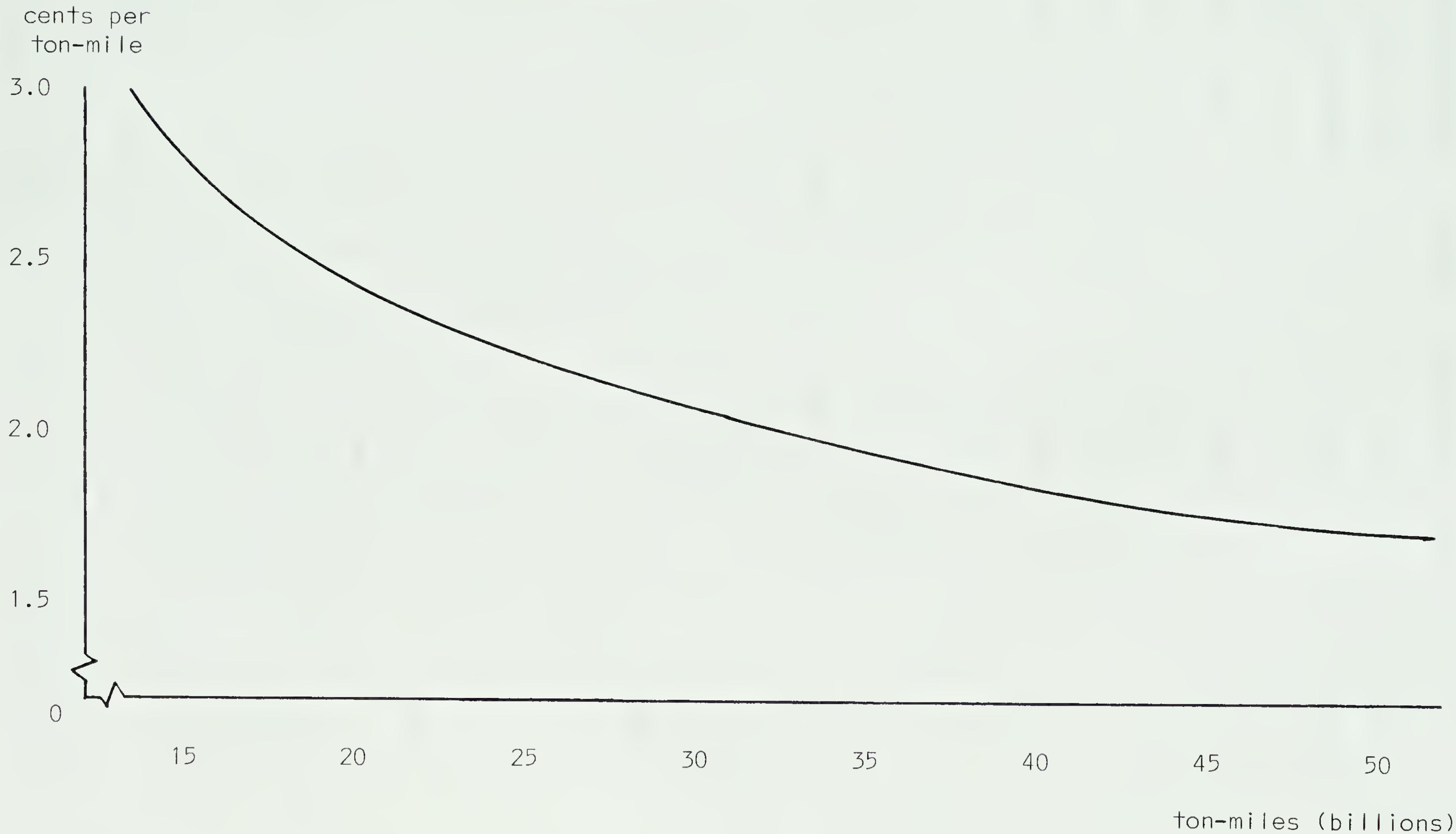


Figure 5-4

INCREMENTAL COSTS
MANUFACTURED GOODS



the calculated values for forest and mining products. Incremental costs increase with output for the years 1952 to 1961 while from 1964 to 1976 constant or slightly declining incremental costs are evident. Upon examination, one observes that the output for forest and mining products from 1952 to 1961 rose by approximately three billion ton-miles or slightly greater than a 25 percent increase in output over nine years. In the next three years, 1961-1964, output rose by over 65 percent, from 14 billion to 23 billion ton-miles and the increase continued until in 1976 it had increased by 135 percent from its 1964 level. This meant that forestry and mining products which had remained at approximately 28 percent of total output from 1952 to 1961 suddenly rose to 32.5 percent of total output in 1964 and up to 41.6 percent by 1976. It would appear that in response to this dramatic increase in demand by this commodity the railways reorganized production of their output and in doing so altered the cost relationship. Output in the manufactured goods class shows a similar increase during the same period with an almost 200 percent increase in output from 1964 to 1976. However, this meant a less dramatic change in the proportion of total output which rose from 33.6 percent in 1961 to 38.7 percent in 1976 and no corresponding change in the cost relation was observed.

Fuss and Waverman (1977) point out that decreasing incremental cost does not necessarily mean in-

creasing returns to scale but any increasing incremental cost curve will indicate decreasing returns. This follows from the fact that a declining incremental cost curve can be the result of a large portion of joint costs being attributed to one output. The joint costs are spread over a greater output as output increases giving the appearance of increasing returns. Rising incremental costs will indicate decreasing returns since if any one output causes total cost to rise by a greater proportion than output, the rise in total cost from all outputs must be greater in proportion than the rise in output. Fuss and Waverman (1976) also note that eventually decreasing returns will exist for any output having $\frac{\partial^2 C}{\partial y_i^2} = \delta_{ii} > 0$. While none of the incremental cost curves can be said to be clearly increasing with output, all four of the δ_{ii}'s are positive indicating at some output level greater than those considered here any of the outputs could lead to overall decreasing returns to density.

5.3 Input and Output Substitution

The Allen partial elasticity of substitution was estimated to examine the input substitution possibilities open to railways. Table 5-5 reports the APES and the price elasticities of factor demand with their standard errors. Values for only 5 of the 25 years are reported as the trends are generally stable. The APES has the expected sign indi-

Table 5-5

ALLEN PARTIAL ELASTICITY OF SUBSTITUTION
and
PRICE ELASTICITIES OF DEMAND[†]

	1953	1958	1964	1970	1976
σ_{LK}*	.583 (.055)	.595 (.053)	.600 (.053)	.594 (.053)	.609 (.051)
E_{LL}*	-.216 (.020)	-.239 (.021)	-.249 (.022)	-.237 (.021)	-.287 (.024)
E_{KK}*	-.367 (.035)	-.356 (.032)	-.350 (.031)	-.357 (.032)	-.325 (.027)
E_{LK}*	.216 (.020)	.239 (.021)	.249 (.022)	.237 (.021)	.284 (.024)
E_{KL}*	.367 (.035)	.356 (.032)	.350 (.031)	.357 (.032)	.325 (.027)

[†]The brackets contain standard errors.

*Significant at the one percent level of confidence.

cating capital and labor are net substitutes as required by production theory for the two input cases. A hypothesis of a constant elasticity of substitution cannot be rejected in this case. The confidence interval for any one of the 25 point estimates will contain all of the remaining point estimates. It should be noted that this result may depend on the simplified input structure specified here and should not be taken as a general result.

Estimates of the price elasticity of factor demand again have the expected signs. It will be noted that $E_{LL} = -E_{LK}$ and $E_{KK} = -E_{KL}$. This result is not generally true but is found here because the production function recognizes only two inputs. The elasticities are assumed to be calculated along an isoproduct curve and, since there are only two inputs, there is only one relative price. A change in either factor price will cause an equal but opposite effect on the price ratio. This in turn leads to an equal but opposite change in the demand as you move along the isoproduct curve. That is, a move from point A to point B on the isoproduct curve may result from a given change in the factor price ratios caused by either an increase in the price of capital for example or a fall in the price of labor.

Oum (1977) points out that as long as the elasticity of demand for the output is negative, the estimated elasticity of factor demand will be greater than the ordinary demand elasticity, ordinary demand elasticity referring to the case where output is allowed to vary. Own

price elasticities will be more elastic when output is allowed to vary while the cross price elasticities would move towards zero with the possibility that they might become negative. Assuming for the moment that there is a single railway output, if it had a price elasticity which was more elastic than -.6, capital and labor would be gross complements rather than gross substitutes. Oum (1977) estimates an elasticity of demand for railway transport and finds a value considerably greater than -.6, which would indicate that capital and labor are both net and gross substitutes.

Allen partial elasticities of transformation were estimated and are reported in Table 5-6 with their standard errors.[11] Five of the six APET have negative signs which indicates the outputs are net substitutes. The one positive elasticity, τ_{AF}, has a standard error approximately three times its own value. The single most striking thing about these elasticities is their magnitudes which are extremely large. Examination of the average revenues per ton-miles and the output levels show that while revenue per ton-mile has remained relatively constant for all four outputs, the level of output for forestry and mining products and manufactured goods both rose very sharply over the study period. This led to the output

---

[11] These non-linear standard errors are calculated by the procedure given by Bodkin and Klein (1967).

Table 5-6

ALLEN PARTIAL ELASTICITIES OF TRANSFORMATION[†]

	1952	1956	1960	1964	1968	1972	1976
τ_{SA}*	-75.88 (25.86)	-85.10 (29.00)	-86.03 (29.31)	-96.88 (33.01)	-101.20 (34.48)	-99.44 (33.86)	-106.46 (36.28)
τ_{SF}*	-28.79 (9.54)	-32.28 (10.70)	-32.63 (10.81)	-36.75 (12.18)	-38.39 (12.72)	-37.72 (12.50)	-40.38 (13.38)
τ_{SM}**	-49.95 (31.03)	-56.02 (34.80)	-56.63 (35.18)	-63.77 (39.62)	-66.62 (41.38)	-65.46 (40.67)	-70.08 (43.53)
τ_{AF}	399.41 (1266.77)	447.95 (1420.74)	452.82 (1436.18)	509.97 (1617.44)	532.70 (1689.53)	523.44 (1660.17)	560.36 (1777.24)
τ_{AM}*	-28.40 (6.86)	-31.85 (7.69)	-32.19 (7.77)	-36.26 (8.75)	-37.87 (9.14)	-37.22 (8.98)	-39.84 (9.62)
τ_{FM}*	-6.46 (1.41)	-7.25 (1.58)	-7.33 (1.60)	-8.25 (1.80)	-8.62 (1.88)	-8.47 (1.85)	-9.07 (1.98)

†The brackets contain standard errors.

*Significant at the one percent level of confidence.

**Significant at the six percent level of confidence, one-tailed test.

ratios changing dramatically for most of the output classes while the rate ratios were stable, which accounts for the large values of the elasticities. The relationships between manufactured goods and agricultural products provides a good example. From 1964 to 1976 the ratio of manufactured goods output to agricultural output rose from 4.26 to 11.48 for a 169 percent increase while the price ratio for the same classes fell by 15 percent from 1.245 to 1.05.

Having discovered the technical reason for the large values of the APET, the question remains as to why rates and output levels behave in this manner. To answer this question note that the railways have faced a problem of excess capacity for a number of decades. This is supported by the increasing returns to density reported earlier in this chapter. With a large amount of excess capacity changes in output levels do not represent movements along the isoproduct curve as required for APET, rather they represent movements from a point within the output set towards the boundary of the set. The observed results are then due to the railways reacting to changing demands for its services and the allocation of the existing excess capacity to meet these increased demands. Since the measures reported here do not appear to measure movements from one equilibrium to another but rather measure a movement from disequilibrium towards equilibrium, they do not represent the transformation possibilities open to the rail-

ways.

Price elasticities of supply are reported in Tables 5-7 to 5-10 with their standard errors. Following the above discussion, the price elasticities also do not measure movements along an isoproduct curve. These elasticities are more a result of changing demand patterns than the production possibilities of the firm. It is not surprising that railways should move quickly to meet any increase in demand considering their excess capacity, what is of interest is the fact that there was no great increase in average revenues per ton-mile in the face of this dramatic increase in demand. Rates, rather than rise in response to the increased demand, fell in real terms during the latter half of the study period. For the manufactured goods output this can be explained by the great increase in competition from trucking at approximately the same time as the demand for this service started to rise. To meet the competition from the trucking industry and capture some of the increased transport demand for themselves, railways were forced to lower rates through agreed charges and other negotiated rates.

This argument does not hold for the forest and mining products output as in general trucking cannot compete in the transportation of low value bulk commodities. To see why the rates for this class have not risen with demand, one must look to the markets in which these products are sold. A great proportion of the goods in the

Table 5-7

ELASTICITIES OF SUPPLY FOR STATUTORY GRAIN†

	1952	1956	1960	1964	1968	1972	1976
$E\delta_{SS}$*	1.64 (.13)	1.08 (.09)	.68 (.06)	1.41 (.11)	.53 (.04)	1.44 (.12)	.61 (.05)
$E\delta_{SA}$*	-5.54 (1.89)	-8.33 (2.84)	-7.00 (2.39)	-6.33 (2.16)	-5.52 (1.88)	-5.23 (1.78)	-4.65 (1.58)
$E\delta_{SF}$*	-6.26 (2.08)	-7.76 (2.57)	-8.05 (2.67)	-9.96 (3.30)	-11.11 (3.68)	-11.70 (3.88)	-15.22 (5.04)
$E\delta_{SM}$**	-26.95 (16.74)	-30.53 (18.97)	-33.12 (20.58)	-34.97 (21.72)	-39.64 (24.63)	-35.70 (22.18)	-37.39 (23.23)

†The brackets contain standard errors.

*Significant at the one percent level of confidence.

**Significant at the six percent level of confidence, one-tailed test.

Table 5-8

ELASTICITIES OF SUPPLY FOR AGRICULTURAL PRODUCTS[†]

	1952	1956	1960	1964	1965	1972	1976
ES_{AA}*	1.00 (.13)	2.19 (.28)	1.27 (.16)	1.33 (.17)	1.18 (.15)	1.28 (.16)	1.25 (.16)
ES_{AS}*	-12.90 (4.40)	-9.93 (3.38)	-7.49 (2.55)	-11.17 (3.81)	-6.17 (2.10)	-9.14 (3.12)	-4.89 (1.67)
ES_{AF}	86.88 (275.56)	107.66 (341.44)	111.69 (354.24)	138.22 (438.38)	154.18 (489.01)	162.31 (514.80)	211.16 (669.71)
ES_{AM}*	-15.32 (3.70)	-17.36 (4.19)	-18.83 (4.55)	-19.88 (4.80)	-22.54 (5.44)	-20.30 (4.90)	-21.26 (5.13)

[†]The brackets contain standard errors.

*Significant at the one percent level of confidence.

Table 5-9

ELASTICITIES OF SUPPLY FOR PRODUCTS OF FORESTS AND MINES†

	1952	1956	1960	1964	1968	1972	1976
ES_{FF}*	.47 (.05)	.67 (.07)	.65 (.07)	1.19 (.13)	1.71 (.19)	2.73 (.30)	11.39 (1.24)
ES_{FS}*	-4.89 (1.62)	-3.77 (1.25)	-2.84 (.94)	-4.24 (1.40)	-2.34 (.78)	-3.47 (1.15)	-1.86 (.62)
ES_{FA}	29.14 (92.42)	43.86 (139.10)	36.85 (116.89)	33.29 (105.60)	29.08 (92.23)	27.54 (87.34)	24.45 (77.55)
ES_{FM}*	-3.49 (.76)	-3.95 (.86)	-4.29 (.94)	-4.53 (.99)	-5.13 (1.12)	-4.62 (1.01)	-4.84 (1.06)

†The brackets contain standard errors

*Significant at the one percent level of confidence.

Table 5-10

ELASTICITIES OF SUPPLY FOR MANUFACTURED GOODS†

	1952	1956	1960	1964	1968	1972	1976
ES_{MM}	1.13 (.14)	1.48 (.19)	1.47 (.18)	2.13 (.27)	3.97 (.50)	4.93 (.62)	9.57 (1.20)
ES_{MS}	-2.07 (.50)	-3.12 (.75)	-2.62 (.63)	-2.37 (.57)	-2.07 (.50)	-1.96 (.47)	-1.74 (.42)
ES_{MA}	-8.49 (5.27)	-6.54 (4.06)	-4.93 (3.06)	-7.35 (4.57)	-4.06 (2.52)	-6.02 (3.74)	-3.22 (2.00)
ES_{MF}	-1.41 (.31)	-1.74 (.38)	-1.81 (.40)	-2.24 (.49)	-2.50 (.55)	-2.63 (.57)	-3.42 (.75)

†The brackets contain standard errors.

*Significant at the one percent level of confidence.

**Significant at the six percent level of confidence, one-tailed test.

forest and mining products category compete on international markets and thus often face very elastic demand curves. This means that railways must keep rates down so that their shippers can compete with other producers. Also contributing to the lack of rate increase for forestry and mining products was the production reorganization which caused the change in the incremental cost curve noted earlier. Technological change is likely to have had a great deal to do with this, since specialized cars, loading facilities and unit trains have greatly changed the nature of the service in this category over the latter half of the study period.

Heaver and Nelson (1977) discuss at length how intermodal and market competition have forced railways to keep rates at or near competitive levels for many shipments. They investigate the post-1967 period, as it was in 1967 that railways were given much more pricing freedom under the National Transportation Act. Increasing returns to density will also have been a contributing factor to the railways' ability to meet the increased demand while keeping rates at competitive levels. It should be noted, however, that given the confidence intervals reported in Table 5-4 an hypothesis of constant returns to density can no longer be rejected. The cost function was also noted to exhibit eventually diminishing returns to density due to the δ_{ii}'s being positive. With this trend towards diminishing returns to density, the railways' future ability to produce

at rates which are competitive will require increased capacity over some sections of track. This study has not been able to estimate returns to plant size therefore the effect of changing the plant size on costs cannot be assessed.

The Snavely and Hall Commissions both document a large amount of excess capacity remains on many lines in western Canada. Given that the measure of average overall returns to density reported in Table 5-4 is tending to constant or diminishing returns, it must be that some lines in Canada are experiencing little or no excess capacity. This being the case, a study using cross-sectional data from these more fully utilized rail sections would provide more accurate estimates of the substitution and transformation possibilities.

CHAPTER SIX

This thesis set out to present a framework based on received production theory that could be useful in discussions of railway costs and regulations. After a brief review of several earlier railway cost studies in Chapter One, Chapter Two proceeded to discuss received production theory as presented by Shephard (1970) and highlight its implications for railway analysis. The translog function and its advantages were discussed in Chapter Three. Among its advantages are the reduced number of maintained hypotheses implied by the function, such as homogeneity of production, separability of production and constant elasticities of substitution and transformation. Chapter Five presented a model of the cost relationships in Canadian railways using the previously discussed techniques. Estimation procedures and results were reported in Chapter Five.

The results in Chapter Five are generally consistent with those of earlier studies. Hypotheses of homogeneity and separability of the production function were both rejected as they were for a re-estimation of Klein's data by Brown *et al.* (1976). Increasing returns to density which were found to be significant for most of the study

period are consistent with the findings of Keeler (1974, 1976). While the incremental cost curves, Allen partial elasticity of substitution and the factor demand elasticities cannot be compared easily to those reported in earlier studies, they seem reasonable in magnitude and of the expected sign. However, when the discussion turns to the output set, results become less reasonable.

The Allen partial elasticities of transformation and the supply elasticities, while having acceptable signs, are very large in magnitude. This would appear to be the result of the observed output adjustments being made within the output set and not along a production possibility boundary. The size of the "elasticities" and the length of time over which the high values persist would suggest that large amounts of excess capacity existed during the early years of the study period. If indeed production is taking place within an output set, then it must be the case that production is taking place in the interior of an input set. This means that the APES and the factor demand elasticities are not along an isoproduct curve as required by their definition, leading to doubt as to their validity. Since other studies have not used such a multiple-output specification, no comparison with earlier work is possible. Noting that it was only in examining the output set that this problem became apparent, one can only speculate as to results if earlier studies had specified multiple outputs.

Unfortunately this means that little can be said in regard to policy direction. A word of caution is warrented, since one should not expect the output-rate relations witnessed in the past decade to necessarily continue as excess capacity diminishes. Railways' actions for the past few decades appear to have been conditioned upon the existence of excess capacity, however upon reaching the boundary of the output set two options will present themselves. First railways may choose to stay in the same output set and react to any changing conditions by substituting along the production possibility boundary. Should this be the chosen course, the magnitude and direction of the output-rate changes need bear no resemblance to those observed while moving within the output set.

A second possibility open to the railways would be to expand capacity so they would move to a larger output set. To understand the effects of the railways' making this choice would require a knowledge of the returns to plant size and the size of an incremental unit of capacity. As noted earlier, returns to plant size could not be investigated in this study, similarly, the aggregated nature of the present data does not allow an examination of where excess capacity has, or is likely to, disappear. It should be noted that it is unlikely that all excess capacity will disappear. Institutional constraints on abandonment of some services such as statutory grain shipments will maintain excess capacity. Due to the peaking of some demands

railways may maintain capacity to meet these occasions thus giving the appearance of excess demand over part of the demand cycle. Any study using less aggregated data will need to consider the above two sources of measured excess capacity.

While the present results from the estimation of the model do not provide a great deal of policy-related conclusions, future studies need not suffer this problem. The framework presented here requires only the proper data to provide results beneficial to discussions of railway policy. Cross-sectional data for different sections of the railways would allow the model to be expanded to include the techniques presented by Keeler (1974, 1976). This would mean that both short-run and long-run cost estimates could be calculated allowing both returns to density and returns to plant size to be investigated.

Considering the great deal of time, energy and money that has been and is continuing to be expended upon the discussion of costs in Canadian railways, it is unfortunate that economic analysis has been restricted to such a minor role in the discussion due to inadequate data. It is hoped that the current effort to demonstrate the possibilities of applying economic analysis to this question will result in further use of such techniques by railways, regulatory agencies and other concerned parties in order to improve the railway system for the benefit of all Canadians.

BIBLIOGRAPHY

Allen, R. G. D. *Mathematical Analysis for Economists.* The University Press, Glasgow: MacMillan, 1938.

Baumol, William J. *Economic Theory and Operation Analysis* (Fourth Ed.). Englewood Cliffs: Prentice-Hall, 1977.

Baumol, William J. and David F. Bradford. "Optimal Departures From Marginal Cost Pricing." *The American Economic Review* 60 (1970): 265-283.

Berndt, Ernst R. and Laurts R. Christensen. "The Internal Structure of Functional Relationships: Separability, Substitution, and Aggregation." *The Review of Economics and Statistics* (March 1973): 403-410.

"The Translog Function and The Substitution of Equipment, Structures and Labor in U.S. Manufacturing 1929-68." *Journal of Econometrics* (March 1973): 81-114.

Bernt, Ernst R. and David O. Wood. "Technology, Prices, and The Derived Demand for Energy." *The Review of Economics and Statistics* (August 1975): 259-268.

Bodkin, Ronald G. and L. R. Klein. "Nonlinear Estimation of Aggregate Production Functions." *Review of Economics and Statistics* (March 1967): 28-44.

Borts, George H. "Production Relations in The Railway Industry." *Econometrica* (1952): 71-79.

"The Estimation of Rail Cost Functions." *Econometrica* (January 1960): 108-131.

Brown, Randall S., Douglas W. Caves and Laurts Christensen. "Estimating Marginal Costs For Multi-product Regulated Firms." Social Systems Research Institute, University of Wisconsin #7609 (July 1976). Mimeograph.

Burgess, David F. "Duality Theory and Pitfalls in the Specifications of Technologies." *Journal of Econometrics* (September 1975): 105-121.

Chipman, J. S. "Returns to Scale in the Railroad Industry: A Reinterpretation of Klein's Data." (abstract) *Econometrica* (1958): 607.

Christensen, Laurts R. and William H. Greene. "Economies of Scale in U.S. Electric Power Generation." *Journal of Political Economy* (August 1976): 655-677.

Christensen, Laurts R., Dale W. Jorgenson and Lawrence Lau. "Conjugate Duality and the Transcendental Logarithmic Function." (abstract) *Econometrica* (July 1971): 255-256.

"Transcendental Logarithmic Production Frontiers." *The Review of Economics and Statistics* (February 1973): 28-45.

Dean, Joel. *Statistical Determination of Costs, with Special Reference to Marginal Cost.* Studies in Business Administration, Vol. 7. University of Chicago Press, 1936.

Diewert, W. E. "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function." *Journal of Political Economy* (May/June 1971):481-507.

Applications of Duality Theory. Department of Manpower and Immigration Canada, 1973.

"Functional Forms for Profit and Transformation Functions." *Journal of Economic Theory* (1973): 284-316.

Fuss, Melvyn and Leonard Waverman. "Multi-product Multi-input Cost Functions For a Regulated Utility: The Case of Telecommunication in Canada." Paper presented at the National Bureau of Economic Research Conference on Public Regulation. Washington, D.C., December 1977.

Gillen, David W. and Michael R. Jones. "Rate Making Principles, Rate Discrimination and The West." *Logistics and Transportation Review* (forthcoming).

Government of Canada. *Grain and Rail in Western Canada,* Vol. 1. The Report of the Grain Handling and Transportation Commission, 1977.

Report of the Commission on the Costs of Transporting Grain by Rail, Vol. 1. 1976.

Griliches, Zvi. "Cost Allocation in Railroad Regulation." *Bell Journal of Economics and Management Sciences* (Spring 1972): 26-41.

Hall, Robert E. "The Specification of Technology with Several Kinds of Output." *Journal of Political Economy* (July/August 1973): 878-892.

Hasenkamp, Georg. *Specification and Estimation of Multiple-Output Production Functions.* New York: Springer-Verlag, 1976.

Heaver, T. D. and James C. Nelson. *Railway Pricing Under Commercial Freedom: The Canadian Experience.* Vancouver: The Centre for Transportation Studies, 1977.

Henderson, James M. and Richard E. Quandt. *Microeconomic Theory: A Mathematical Approach,* 2nd Ed. New York: McGraw-Hill Book Co, 1971.

Keeler, Theodore E. "Railroad Costs, Returns to Scale, and Excess Capacity." *The Review of Economics and Statistics* (May 1974): 201-208.

"On the Economic Impact of Railroad Freight Regulation." Berkeley: University of California, Working Paper No. SL-7601, September 1976.

Kneafsey, James T. *Transportation Economic Analysis.* Toronto: Lexington Books, 1975.

Nikaido, Hukukane. *Convex Structures and Economic Theory.* New York: Academic Press, 1968.

Oum, Tae Hoon. "Derived Demand for Freight Transportation and Substitutabilities Among Modes in Canada." Centre for Transportation Studies, University of British Columbia, 1977. (Mimeograph).

Purdy, H. L. *Transport Competition and Public Policy in Canada.* Vancouver: University of British Columbia Press, 1972.

Shephard, Ronald W. *Theory of Cost and Production Functions.* Princeton, New Jersey: Princeton University Press, 1970.

Spady, R. H. and A. F. Freidlander. "Hedonic Cost Functions for the Regulated Trucking Industry." *The Bell Journal of Economics* (Spring 1978): 159-179.

Transportation. A submission by the four western provinces to the Western Economic Opportunities Conference, July 24-26, 1973, Calgary.

Water, W. G. II. "Statistical Costing in Transportation." *Transportation Journal* (Spring 1976): 49-62.

Westin, R. "A Reduced Form Cost Function Approach to the Specification of Freight Demand Models." University of Toronto, March 1977. Unpublished mimeograph.

Uzawa, Hirofumi. "Duality Principles in the Theory of Cost and Production." *International Economic Review* (May 1964): 216-220.

APPENDIX A

To simplify the notation for ease of exposition, rewrite the production function (4.4) as

(A.1) $$f(y_1, \ldots y_m, x_{m+1}, \ldots, x_{m+n}) = 0$$

where the y's are outputs and the x's are inputs, measured negatively. The APET can now be rewritten as

(A.2) $$\tau_{ij} = \frac{y_1 f_1 + y_2 f_2 + \ldots + y_m f_m}{y_i y_j} \frac{D_{ij}}{D} \qquad i,j = 1, \ldots, m \quad i \neq j$$

$$\text{where } D = \begin{vmatrix} 0 & f_1 & f_2 & f_3 & \ldots & f_m \\ f_1 & f_{11} & f_{12} & f_{13} & & . \\ f_2 & f_{21} & f_{22} & f_{23} & & . \\ . & & & & & . \\ . & & & & & \\ . & & & & & \\ f_m & . & . & . & & f_{mm} \end{vmatrix}$$

and D_{ij} is the co-factor of the term f_{ij} of D. Defining the term K_i to be the revenue share of the output y_i,

(A.3) $$K_i = \frac{y_i q_i}{R} = \frac{y_i q_i}{y_1 q_1 + y_2 q_2 + \ldots + y_m q_m} \qquad i = 1, \ldots, m$$

where q_i is the price of y_i and R is total revenue. This can be rewritten using the result that in equilibrium f_i is proportional to the price of y_i.

(A.4) $$K_i = \frac{y_i f_i}{y_1 f_1 + y_2 f_2 + \ldots + y_m f_m} \qquad i = 1, \ldots, m$$

The first order conditions for profit maximization require that

(A.5) $$q_i = - p_{m+j} \frac{f_i}{f_{m+j}} \qquad i = 1, \ldots, m \quad j = 1, \ldots, n$$

where p's are the prices of the inputs. Setting j = 1 and taking the

partial derivatives of the production function (A.1) and the m price equations (A.5) with respect to q_1 the following system is generated when outputs are held constant.

$$f_1 \frac{\partial y_1}{\partial q_1} + f_2 \frac{\partial y_2}{\partial q_1} + \ldots + f_m \frac{\partial y_m}{\partial q_1} = 0$$

$$-\frac{\partial p_{m+1}}{\partial q_1}\frac{f_1}{f_{m+1}} - \frac{p_{m+1}}{f_{m+1}}\left[f_{11}\frac{\partial y_1}{\partial q_1} + f_{12}\frac{\partial y_2}{\partial q_1} + \ldots + f_{1m}\frac{\partial y_m}{\partial q_1}\right] = 1$$

$$-\frac{\partial p_{m+1}}{\partial q_1}\frac{f_2}{f_{m+1}} - \frac{p_{m+1}}{f_{m+1}}\left[f_{21}\frac{\partial y_1}{\partial q_1} + f_{22}\frac{\partial y_2}{\partial q_1} + \ldots + f_{2m}\frac{\partial y_m}{\partial q_1}\right] = 0$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$-\frac{\partial p_{m+1}}{\partial q_1}\frac{f_m}{f_{m+1}} - \frac{p_{m+1}}{f_{m+1}}\left[f_{m1}\frac{\partial y_1}{\partial q_1} + f_{m2}\frac{\partial y_2}{\partial q_1} + \ldots + f_{mm}\frac{\partial y_m}{\partial q_1}\right] = 0$$

Rearranging terms this system can be written as

$$f_1 \frac{\partial y_1}{\partial q_1} + f_2 \frac{\partial y_2}{\partial q_1} + \ldots + f_m \frac{\partial y_m}{\partial q_1} = 0$$

$$f_1\left(\frac{1}{p_{m+1}}\frac{\partial p_{m+1}}{\partial q_1}\right) + f_{11}\frac{\partial y_1}{\partial q_1} + f_{12}\frac{\partial y_2}{\partial q_1} + \ldots + f_{1m}\frac{\partial y_m}{\partial q_1} = -\frac{f_{m+1}}{p_{m+1}}$$

$$f_2\left(\frac{1}{p_{m+1}}\frac{\partial p_{m+1}}{\partial q_1}\right) + f_{21}\frac{\partial y_1}{\partial q_1} + f_{22}\frac{\partial y_2}{\partial q_1} + \ldots + f_{2m}\frac{\partial y_m}{\partial q_1} = 0$$

$$f_m\left(\frac{1}{p_{m+1}}\frac{\partial p_{m+1}}{q_1}\right) + f_{m1}\frac{\partial y_1}{\partial q_1} + f_{m2}\frac{\partial y_2}{\partial q_1} + \ldots + f_{mm}\frac{\partial y_m}{\partial q_1} = 0$$

Using Cramer's rule to solve for $\frac{\partial y_1}{\partial q_1}$ when the x's are held constant.

$$\text{(A.6)} \qquad \frac{\partial y_1}{\partial q_1} = -\frac{f_{m+1}}{p_{m+1}}\begin{vmatrix} 0 & f_2 & f_3 & \cdots & f_m \\ f_2 & f_{22} & f_{23} & & \cdot \\ f_3 & f_{32} & f_{33} & & \cdot \\ \vdots & & & & \cdot \\ f_m & & \cdots & & f_{mm} \end{vmatrix} \div D = -\frac{f_{m+1}}{p_{m+1}}\frac{D_{11}}{D}$$

Substituting (A.5) into (A.6)

(A.7) $$\frac{\partial y_1}{\partial q_1} = \frac{f_1}{q_1} \frac{D_{11}}{D}$$

Multiplying both sides of (A.7) by $\frac{q_1}{y_1}$

(A.8) $$\frac{\partial y_1}{\partial q_1} \frac{q_1}{y_1} = \frac{f_1 q_1}{y_1 q_1} \frac{D_{11}}{D} = \frac{f_1 y_1}{y_1 y_1} \frac{D_{11}}{D}$$

This expression can be rewritten with the substitution of (A.4)

(A.9) $$\frac{\partial y_1}{\partial q_1} \frac{q_1}{y_1} = \frac{K_1 (y_1 f_1 + y_2 f_2 + \dots + y_m f_m)}{y_1 y_1} \frac{D_{11}}{D} = K_1 \tau_{11}$$

Following the same procedure each of the m x m partial derivatives can be found. One further result will indicate the generalization, therefore solving for $\frac{\partial y_2}{\partial q_1}$

(A.10) $$\frac{\partial y_2}{\partial q_1} = - \frac{f_{m+1}}{p_{m+1}} \begin{vmatrix} 0 & f_1 & f_3 & \dots & f_m \\ f_2 & f_{21} & f_{23} & & \cdot \\ f_3 & f_{31} & f_{33} & & \cdot \\ \vdots & & & & \cdot \\ f_m & \cdot & \cdot & \cdot & f_{mm} \end{vmatrix} \div D$$

$$= - \frac{f_{m+1}}{p_{m+1}} \frac{D_{12}}{D}$$

Using similar substitution, the following results.

(A.11) $$\frac{\partial y_2}{\partial q_1} \frac{q_1}{y_2} = \frac{K_1 (y_1 f_1 + y_2 f_2 + \dots + y_m f_m)}{y_1 y_2} \frac{D_{12}}{D} = K_1 \tau_{12}$$

Generalizing the above results

$$\frac{\partial y_i}{\partial q_j}\frac{q_j}{y_i} = K_j\tau_{ij} \qquad i,j = 1, \ldots, m \tag{A.12}$$

The expression on the left of (A.12) corresponds to the definition of the price elasticity of supply (ES). The price elasticity of supply is directly related to the APET through the revenue share.

$$ES_{ij} = K_j\tau_{ij} \qquad i,j = 1, \ldots, m \tag{A.13}$$

By substituting (A.3) into (A.12) a new expression for τ_{ij} can be derived.

$$\frac{\partial y_i}{\partial q_j}\frac{q_j}{y_i} = \frac{y_j q_j}{R}\tau_{ij} \tag{A.14}$$

$$\tau_{ij} = \frac{R\frac{\partial y_i}{\partial q_j}}{y_i y_j} \qquad i,j = 1, \ldots, m$$

To put τ_{ij} in terms of a cost function, first recall from Shepahrd's lemma

$$\frac{\partial C}{\partial y_j} = q_j$$

then
$$\frac{\partial}{\partial y_i}\left(\frac{\partial C}{\partial y_j}\right) = \frac{\partial}{\partial y_i}(q_j)$$

$$\frac{\partial^2 C}{\partial y_i \partial y_j} = \frac{\partial q_j}{\partial y_i} \tag{A.15}$$

The inverse function rule states that

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}$$

Equation (A.15) and the inverse function rule provide the following

(A.16) $$\frac{\partial y_i}{\partial q_j} = \frac{1}{\frac{\partial q_j}{\partial y_i}} = \frac{1}{\frac{\partial^2 C}{\partial y_i \partial y_j}}$$

Substituting (A.16) into (A.14)

(A.17) $$\tau_{ij} = \frac{R}{y_i y_j \frac{\partial^2 C}{\partial y_i \partial y_j}} \qquad i,j = 1, \ldots, m$$

In order to express τ_{ij} in terms of the translog cost function it is necessary to evaluate $\frac{\partial^2 C}{\partial y_i \partial y_j}$. First take the partial derivative of $\ln C$ with respect to $\ln y_i$.

(A.18) $$\frac{\partial \ln C}{\partial \ln y_i} = \beta_i + \sum_j^m \delta_{ij} \ln y_j + \sum_j^n \rho_{ij} \ln p_j \qquad i = 1, \ldots, m$$

This expression can be written as

(A.19) $$\frac{\partial C}{\partial y_i} = \frac{C}{y_i}\left[\beta_i + \sum_j^m \delta_{ij} \ln y_j + \sum_j^n \rho_{ij} \ln p_j\right] \qquad i = 1, \ldots, m$$

Taking the partial derivative of (A.19) with respect to y_j will give the desired result. It must be noted, however, that there are two cases, first when $i = j$ and second when $i \neq j$. First evaluating the case where $i \neq j$

(A.20) $$\frac{\partial^2 C}{\partial y_i \partial y_j} = \frac{C\delta_{ij}}{y_i y_j} \qquad i,j = 1, \ldots, m \quad i \neq j$$

When $i = j$ the result is

(A.21) $$\frac{\partial^2 C}{\partial y_i \partial y_i} = \frac{C\delta_{ii}(1 - \ln y_i)}{y_i^{\,2}} \qquad i = 1, \ldots, m$$

Substituting (A.20) and (A.21) into (A.17) the translog forms of τ_{ij} are

(A.22) $$\tau_{ii} = \frac{R}{C\delta_{ii}(1 - \ln y_i)} \qquad i = 1, \ldots, m$$

(A.23) $$\tau_{ij} = \frac{R}{C\delta_{ij}} \qquad i,j = 1, \ldots, m \quad i \neq j$$